UNVERSITY OF ALBERTA

Development and characterization of microsatellite DNA markers in *Picea glauca* (Moench) Voss. (white spruce)

By

Heather Christine Cobban ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

DEPARTMENT OF FOREST SCIENCE

EDMONTON, ALBERTA

FALL 1997

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22587-9

Canada

University of Alberta

Library Release Form

Name of Author: **Heather Christine Cobban**

Title of Thesis: **Development and characterization of microsatellite DNA markers in *Picea glauca* (Moench) Voss. (white spruce)**

Degree: **Master of Science**

Year this Degree Granted: **1997**

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Heather Cobban

214, 2433 Malaview Avenue

Sidney, B C V8M 1K1

Sept. 8, 1997

date submitted to the Faculty of

Graduate Studies and Research

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE DNA MARKERS IN *Picea glauca* (Moench) Voss. (WHITE SPRUCE)** submitted by **HEATHER C. COBBAN** in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE**.

Dr. B. P. Dancik

Supervisor

Dr. A. G. Good

Co-supervisor

Dr. C. Strobeck

Dr. S. E. Macdonald

DATE: Sept. 8, 1997

Abstract

White spruce [*Picea glauca* (Moench) Voss] is the most extensively planted forest tree species in Canada. Yet, to date, few molecular data have been gathered to determine the natural patterns and distribution of genetic diversity of this species. Microsatellite DNA markers, which require only minute amounts of sample DNA, and are efficient to process, are well suited to collection of such data. Because of their hyper-variability, neutrality, co-dominance and ubiquity in many eukaryotic genomes, microsatellites are gaining preference as molecular markers for population genetic studies, as well as for strain identification, paternity analysis, and mapping purposes. I isolated and characterized 10 microsatellites from the white spruce genome, and designed primers for seven of them, averaging 18 consecutive repeats. Amplification patterns for these loci contrasted with those characterized in most other species. Of the seven loci assessed using the polymerase chain reaction, none amplified as single-locus, polymorphic markers, indicating that microsatellite loci in white spruce may be less informative than in other species. Four microsatellites amplified as unique, but monomorphic loci. Three microsatellites amplified as multi-locus (i.e., more than two bands per individual) and polymorphic. One of these multi-locus microsatellites failed to amplify the target sequence. Altogether, six of ten microsatellites isolated were associated with repetitive DNA. This contrasts with estimates of 25% of microsatellites associated with repetitive DNA in other forest tree species. Frequency of the $(CT)_n$ repeat in the white spruce genome was estimated at one repeat in excess of 11 units per 704 kb of genome, which is considerably lower than has been reported in most other genomes to date, with the exception of *Pinus radiata*.

Acknowledgments

I thank my supervisor, Dr. Bruce Dancik, for his generous support and assistance throughout this project. I also thank my co-supervisor, Dr. Allen Good, for his generosity in sharing both lab space and materials.

Dr. P. Khasa's significant contribution of material support throughout this project was most appreciated. Thanks also to Dr. Curt Strobeck and his lab for sharing both protocols and computers. I am particularly grateful to Gary Ritzel for providing valuable technical assistance during several complex stages of lab work. Dr. O. Rajora assisted with white spruce sample collections, and arranged for additional assistance through Weyerhaeuser in Prince Albert, and the Prince Albert Model Forest in Saskatchewan. Roman Orynik, of Weyerhaeuser, in particular, facilitated our sample collection process. Others who assisted in many ways include fellow lab workers in BSG404, and the fourth floor Biosciences "Lunch Club".

Funding for this project was contributed by NSERC, through Dr. B. P. Dancik and Dr. P. Khasa, by the Prince Albert Model Forest, through Dr. O. Rajora, and by a Max MacLaggan Scholarship.

Table of Contents

Chapter 1

Introduction

1.1 Objectives 1

1.2 General Background Information on Microsatellite Repeats 1

1.2.1 Rationale for Isolating in the White Spruce Genome 3

1.2.2 Evolutionary Distribution of Microsatellites 14

1.2.3 Assessment of Polymorphism at Microsatellite Loci 16

1.2.4 Classification of Microsatellites 17

1.2.5 Informativeness of Microsatellites 18

1.2.6 Microsatellite frequencies vary with repeat number 20

1.2.7 Distribution of Microsatellites within Plant Genomes 20

1.2.8 Mutational Mechanism of Microsatellites 23

1.2.9 Inheritance of Microsatellite Alleles 27

1.2.10 Distribution of Microsatellite Alleles 28

1.2.11 Polymorphism Associated with Plant Microsatellites 29

1.2.12 Potential Uses of Microsatellites 29

Chapter 2

Materials and Methods 45

2.1 Population Sampling and Experimental Design 45

2.2 DNA extraction from White Spruce Population Samples 45

2.3 Genomic Library Construction 46

2.3.1 DNA Extraction 46

2.3.2 Preparation of Genomic Library Inserts 46

2.3.3 Preparation of Vector for Ligation with Library Inserts 47

2.3.4 Ligation of White Spruce Genomic Inserts into pBluescript 47

2.3.5 Preparation of $CaCl_2$ Competent Cells 48

2.3.6 Transformation of Ligated Inserts into XL1-Blue.MRF' 48

2.4 Screening of Genomic Library for Microsatellite Repeats 50

2.4.1 Colony Lifts of Genomic Library 50

2.4.2 End-Labeling of Oligonucleotide Probes 50

2.4.3 Prehybridization of Colony Lifts 51

2.4.4 Hybridization of Colony Lifts to Oligonucleotide Probes 51

2.4.5 Identification of Positive Colonies from Genomic Library 52

2.5 Preparation of Template for Commercial Cycle Sequencing 52

2.6 Design of Primers Flanking Microsatellite Sequences 53

2.7 PCR Assessment of Microsatellite Primers on White Spruce Population Samples 54

2.8 Estimate of Quantity of White Spruce Genome Screened 55

Chapter 3

Results 57

3.1 Frequency and Composition of $(CT)_n$ Microsatellites in White Spruce 57

3.2 Informativeness of White Spruce Microsatellites 57

Chapter 4

Discussion 69

4.1 Characterization of Microsatellites in White Spruce 69

4.2 Summary 72

List of Tables

Table 3.1	White Spruce Microsatellites	58
Table 3.2	White Spruce Microsatellite Primers	59
Table 3.3	White Spruce Microsatellite Amplification Patterns	68

List of Figures

Figure 1.1	The Botanical Range of White Spruce	4
Figure 1.2	Identifying Features of White Spruce	5
Figure 1.3	Mechanisms of Mutation within Tandem Repeat Sequences.	24
Figure 3.1	Amplification of Locus 3.4.b within Snowfield Road Old Growth Population Samples	60
Figure 3.2	Amplification of Locus 7.2.a within Montreal Lake Natural Regeneration Population Samples	62
Figure 3.3	Amplification of Locus 7.2.a within Montreal Lake Planted Population Samples	63
Figure 3.4	Amplification of Locus 7.2.a within Montreal Lake Old Growth Population Samples	64
Figure 3.5	Amplification of Locus 7.2.a within Snowfield Road Natural Regeneration Population Samples	65
Figure 3.6	Amplification of Locus 7.2.a within Snowfield Road Planted Population Samples	66
Figure 3.7	Amplification of Locus 7.2.a within Snowfield Road Old Growth Population Samples	67

Chapter 1 Introduction

1.1 Objectives:

This project encompassed two main objectives:

(1) To identify and characterize microsatellite repeats in the white spruce [*Picea glauca* (Moench) Voss] genome; and

(2) To utilize microsatellite repeats to assess genetic diversity in populations of natural versus planted white spruce, in order to determine whether management of white spruce populations has altered the genetic diversity of this species.

1.2 General Background Information on Microsatellite Repeats

Microsatellite repeats (Litt and Luty, 1989), also known as simple sequence repeats (SSR's) (Jacob, Lindpainter, Lincoln, Kusumi, Bunker, Mao, Yi-Pei, Ganten, Dzau, and Lander, 1991), sequence tagged microsatellite sites (STMS's) (Beckman and Soller, 1990), and short tandem repeats (STR's) (Edwards, Civitello, Hammond and Caskey, 1991) are tandemly occurring repeats of very short motifs, consisting of one to five nucleotides, such as, for example, $(CA)_n$, or $(CT)_n$, where 'n' refers to the number of repeats at

any given locus. In this thesis, I will use the acronym SSR, in addition to the term "microsatellite".

Microsatellites have become the marker of choice in a number of population genetic studies for several reasons. Firstly, they have been found to be much more polymorphic than other single locus marker systems used to date in many higher organisms, including *Homo sapiens* (Litt and Luty, 1989; Weber and May, 1989; Tautz, 1989), *Arabidopsis* (Bell and Ecker, 1993), *Oryza* species (rice) (Wu and Tanksley, 1993), *Glycine* species (soybean) (Morgante and Olivieri, 1994; Rongwen and Akkaya, 1995), *Pinus radiata* (Monterey pine) (Smith and Devey, 1994), *Dioscorea tokoro* (wild yam) (Terauchi and Konuma, 1994), *Quercus macrocarpa* (bur oak) (Dow, Ashley and Howe, 1995), *Citrus* (citrus) and related species (Kijas, Fowler and Thomas, 1995), *Pinus strobus* (white pine) (Echt, May-Marquardt, Hseih and Zahorchak, 1996), *Triticum aestivum* (wheat) (Ma, Roder and Sorrells, 1996), *Solanum tuberosum* (cultivated potato) (Provan, Powell and Waugh, 1996), *Vitis vinifera* (grape) (Thomas and Scott, 1993; Bowers, Dangl, Vignani and Meredith, 1996), and *Zea mays* (maize) (Taramino and Tingey, 1996). In addition, several microsatellites in vertebrates and the few plant species studied to date have been found to be both evenly and abundantly distributed throughout the genomes, and apparently selectively neutral (Weber, 1990; Wu and Tanksley, 1993; Hite, Eckert and Cheng, 1996; Taramino and Tingey, 1996). Analysis of microsatellite loci is facilitated by the co-dominant nature of markers generated, allowing heterozygotes to be distinguished from homozygotes (Queller, Strassman and Hughes, 1993). While microsatellites are considerably more expensive to develop initially than other marker

systems, once identified, they offer a relatively quick and cost effective means of analyzing large numbers of samples (Queller *et al*, 1993; Rafalski and Tingey, 1993). They more easily lend themselves to automated processing than, for example, either restriction fragment length polymorphisms or isozymes (Rafalski and Tingey, 1993). Once published, SSR primer sequences are immediately available to other researchers working on the same or similar species, in contrast to, for example, RFLP probes, which must be physically exchanged between labs (Rafalski and Tingey, 1993).

1.2.1 Rationale for Isolating Microsatellites in the White Spruce Genome

White spruce [*Picea glauca* (Moench) Voss] represents the most extensively planted forest tree species in Canada, comprising 35.2% of all trees planted in Canada during the period from 1980 to 1986 (D. H. Kuhnke, 1989) (see Figures 1.1 and 1.2). The area of Canada's forest harvested during an eleven year period, from 1975 to 1986, increased by 32.1%, while the area planted during this same period more than doubled. These statistics indicate trends toward both more intensive and extensive management of Canadian forests in the future.

In order to preserve levels and patterns of genetic diversity present in natural populations of white spruce, it will be necessary to establish baseline data on the genetic diversity of this species, before large-scale intervention in its natural growth alters such patterns. Detailed baselines of genetic diversity would allow potential impacts of modern forest management activities upon the genetic structure of current and future

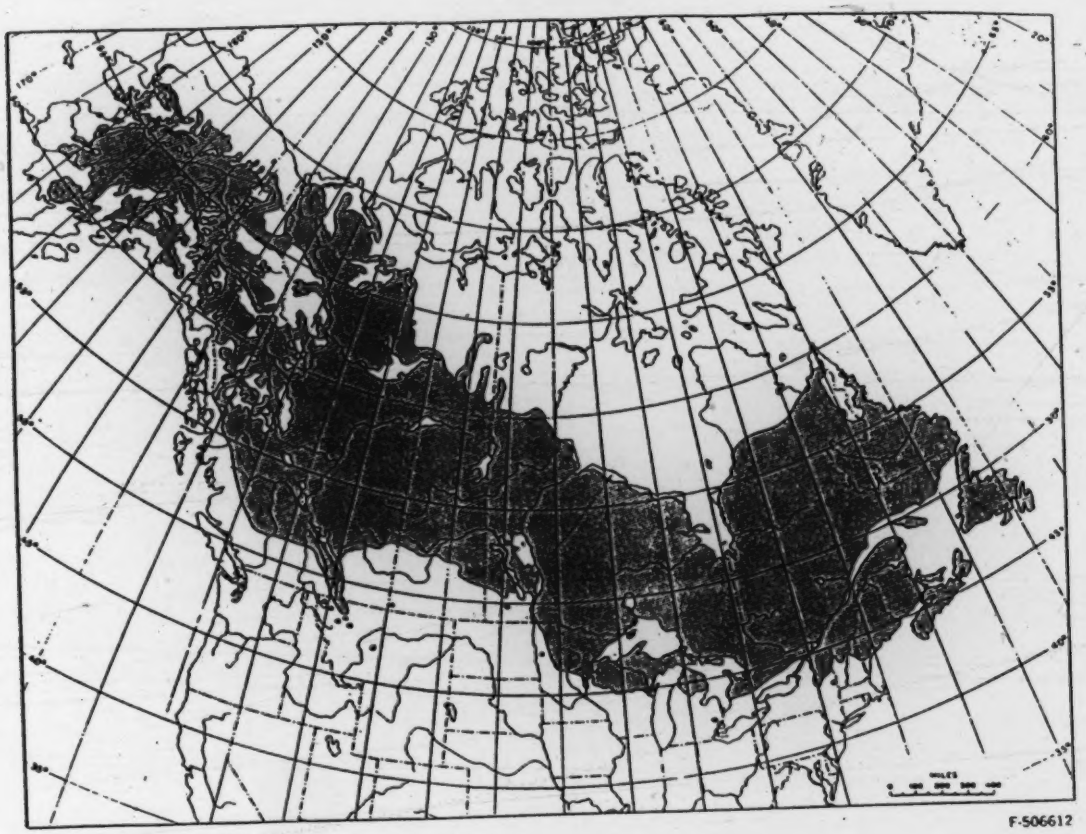

Figure 1.1. The botanical range of white spruce (Fowells, 1965)

Figure 1.2 Identifying features of white spruce (Preston, 1989).
(a) branchlet with cone, (x 1); (b) seeds, (x 1)

forests to be monitored (Allard, 1970; Frankel, 1974; Marshall and Brown, 1975; Brown, 1978; Adams, 1981; Eldridge, 1990; Ledig, 1992).

Several approaches of modern forest management have the potential to alter the genetic base of boreal forest tree species. Large-scale clear-cutting may reduce genetic diversity by physically removing whole populations of trees (Adams, 1981; Brown, 1992; Ledig, 1992). If some of these populations represent locally adapted ecotypes, their genotypes may not be adequately represented elsewhere (Anderson, 1949; Bradshaw, 1972; Kleinschmidt, 1979). Clinal patterns of variation have been observed for many tree species, including Ponderosa pine (*Pinus ponderosa* Dougl. ex Laws.) (Conkle, 1973), white spruce (Alden and Loopstra, 1987) and Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] (Campbell, 1979). While large-scale deforestation events, such as the permanent loss of Mediterranean forests (Hughes, 1982), or the conversion of Scottish Caledonian pine forests to heathland (Carlisle, 1977), are well documented, many more localized genetic losses go unrecorded due to lack of sufficient baseline data (Ledig, 1992). Information pertaining to the structure of genetic diversity of white spruce across the landscape would be useful in planning for both protection and harvest activities of this species.

A second aspect of forest management with much potential to alter the genetic structure of forest tree species involves artificial regeneration (Cleary, Greaves and Hermann, 1978; Silen and Doig, 1976; Silen and Wheat, 1979; Adams, 1981; Brown, 1992; Ledig, 1992). In the absence of genetic baseline information for forest tree species, seed transfer guidelines have been established based upon predictions of the spatial distribution of

genetic diversity for a species, rather than on measurements of actual distribution (Li, Beaulieu and Bousquet, 1993; Ledig, 1992). While forest species have, in some cases, been shown to diverge over short distances (Fryer and Ledig, 1972; Campbell, 1979), seed zones are generally established across larger scales (Ledig, 1992). Planted seedlings may also originate from a more restricted genetic base than their natural counterparts (Adams, 1981; Ledig, 1992). It is therefore possible that seedlings used to re-stock areas may differ significantly in composition from the natural populations which they replace. Should family structure be influencial in the mating system of a species (Wright, 1953; Coles and Fowler, 1976; Cheliak, Pitel and Murray, 1984; King, Dancik and Dhir, 1984), this is also likely to be altered by artificial reforestation (Ledig, 1992). Once outplanted, artificially regenerated stock differing substantially from the original tree population could have long-term genetic effects, both on the plantation site and beyond it, through hybridization with surrounding natural populations, followed by introgression of foreign genes into natural stands (Brown, 1992; Ledig, 1992). If such genes are not of long-term adaptive value, consequences to local populations could be both detrimental and costly. Gene migration from planted trees into natural populations has been observed in two of three endemic populations of Monterey pine in California, as well as one endemic population of Bishop pine (*Pinus Muricata* D. Don), in California, with as yet unknown long-term consequences (Millar and Libby, 1989).

Based on Canadian forestry statistics from 1983 to 1986, clear-cut harvesting represented 89.6% of all harvesting in Canada (D. H. Kuhnke, 1989). Between 1982 and 1986, 31% of harvested areas were artificially

regenerated, with an approximate success rate of 65%, resulting in successful re-stocking of 20% of harvest sites. Among artificially regenerated seedlings reported to have survived to the establishment phase during the above period, a lower percentage is expected to survive to maturity, suggesting that second growth stands will be less productive than the natural sites which they replace (Drew, 1987; Frisque, Weetman, and Clemmer, 1978). Surveys by Drew (1987) found up to 38% of cutover areas previously declared to be satisfactorily stocked to be no longer satisfactorily stocked within 5-10 years. Kuhnke (1989) likewise reported a subsequent reversion of almost 32% of cutover areas to an inadequately forested state.

Many generations of selection through both seasonal environmental stresses and rare environmental extremes have conditioned natural stands for long-term survival on their specific growth sites (Ferrell and Woodard, 1966; Callaham, 1970; Bradshaw, 1972; Roche, 1975; Wright, 1976; Campbell, 1979; Kleinschmidt, 1979; Adams, 1981; Kimmins, 1987; Brown, 1992; Ledig, 1992). Based on a five year study of seedfall within natural stands of Douglas-fir, Campbell (1979) estimated that only one of 20,000 seeds produced would survive to become an old growth tree. Of course, this high rate of exclusion is not entirely due to natural selection, since chance events, independent of genetics, will bar many seeds from germinating. Still, among seedlings which germinate, only those which are able to compete with surrounding vegetation and cope with natural environmental stresses will survive to maturity (Kimmins, 1987). The presence of multiple cohorts within natural stands increases genetic diversity, since different cohorts may be subjected to varying selective

regimes during their development (Mulcahy, 1975; King, Dancik and Dhir, 1984). This enhanced diversity increases the odds that in any given year, some seed will be well adapted to the growing conditions (Kimmins, 1987). Nursery grown seedlings, in contrast, are germinated and grown under optimal conditions, being highly sheltered from selection until placed in the field, perhaps on sites for which they are not ideally suited (Adams, 1981). In some cases, plantation failure due to poorly adapted seedling sources may occur years or decades after planting, representing significant economic losses (Silen, 1978; Kimmins, 1987). Silen (1978) reported on a 1912 study in which Douglas-fir seedlings from 13 locations, representing 120 families, were planted across five differing elevations. Severe mortality of low elevation seedlings planted at the highest elevation became evident within the first decade. On less exposed sites, however, decimation of non-adapted seed sources did not occur until three decades after planting. While it is possible that non-local seed sources could perform favourably on divergent sites (Kimmins, 1987), this study illustrates the need for long-term testing to adequately determine this. It is possible that marker-assisted selection could assist in predicting growth performance of seedlings under varying field conditions, as well as in identifying sources of seedling stocks, in order to facilitate optimal assignment of stocks to field locations (Nienstaedt and Teich, 1972).

The cultivation process, itself, is noted for reducing genetic diversity within domesticated species, due to the uniformity in growth characteristics favoured by mechanized and artificially structured nursery regimes (Kimmins, 1987; Brown, 1992; Ledig, 1992). Such artificial growth regimes impose many inadvertent selective forces on natural stocks, which

might otherwise perform well in the field (Campbell and Sorensen, 1984; El-Kassaby, 1992; Theisen, 1980, Kitzmiller, 1990; Stoehr and El-Kassaby, 1997). Seed which fails to germinate on time because of insufficient stratification, for example, may be culled due to delayed growth (El-Kassaby, 1992; Ledig, 1992; Stoehr and El-Kassaby, 1997). Likewise, seedlings which fail to develop sufficient root structure prior to lifting date (Jenkinson, 1984), or which experience transplantation shock (Beineke and Perry, 1966), may not survive outplanting, even on sites for which they would be well adapted. Among many studies which have documented loss of genetic diversity through cultivation, two recent studies compared levels of microsatellite polymorphism in wild and cultivated species of rice and soybean, and found reduced levels of diversity in both cultivated species. Observed heterozygosity in rice fell from 0.79 in wild species to 0.57 in cultivated lines (Wu and Tanksley, 1993), while wild soybean exhibited 84% genetic differentiation, in contrast to 43% differentiation within elite germplasm (Morgante, Rafalski, Biddle, Tingey and Olivieri, 1994).

Many forest tree species are among the most genetically diverse of all higher organisms due to their longevity, extensive ranges, high fecundity, later successional habitat, and open-pollinated, outcrossing mating systems (Hamrick and Godt, 1979; Hamrick, Linhart and Mitton, 1979). Isozyme studies have confirmed high levels of genetic diversity in populations of white spruce, congruent with diversity levels of most other conifers studied to date (King, Dancik and Dhir, 1984; Cheliak, Pitel and Murray, 1985).

Because many white spruce populations currently exist in a relatively natural state, there is an opportunity to conserve high levels of genetic diversity for future stands of this species, through careful monitoring and planning. Such opportunities have been lost for many other domesticated plant species, such as maize, which has been rendered vulnerable to massive disease infestations by its drastically reduced genetic base (Plucknett, Smith, Williams and Murthi-Anishetti, 1987; Smith, 1988; Levings, 1990), and for which the wild ancestors are believed to have been completely eradicated and replaced with more modern cultivars (Mangelsdorf, MacNeish and Galinat, 1964; Mangelsdorf, 1974 ; Ledig, 1992).

Most domesticated plant species have additionally experienced sharp declines in genetic diversity as a result of modern plant breeding methods, which both reduce genetic diversity through persistent directional selection, and distribute successful cultivars globally, eliminating diversity which might otherwise arise through divergent local breeding programs (Day, 1973; Frankel, 1974; Harlan, 1972; Marshall, 1977; Barrett, 1981; Brown, 1992). Throughout history, humans have derived food from approximately 3000 plant species (Mangelsdorf, 1966; Harlan, 1975). Because some species were better suited to domestication, and in the interests, presumably, of mass production, a trend ensued of planting larger areas with fewer cultivars, to the point where we now rely principally on the specialized cultivars of only 15 plant species to supply 78% of the world's food, as measured in gross tonnage (Barrett, 1981). The replacement of wild stocks with increasingly specialized cultivars has caused an erosion of genetic resources, which in many cases will limit future breeding programs for these species. Should the need arise to introduce genes for new uses, or to

cope with unforeseen stresses in the future, the originally diverse wild stocks will no longer be there as sources of new diversity.

Within those few species so extensively planted today, seed stocks originate from increasingly restricted sources. In several American states, for example, single cultivars have comprised 99% of planted wheat (Day, 1973). It is not uncommon today for elite cultivars to be planted across entire continents (Barrett, 1981). Modern plant breeding is likely to play a greater role in regenerated stands of conifers in the future, as pressure increases to produce ever more timber and fiber on increasingly restricted areas (Zobel, 1974; Silen and Wheat, 1979). In "The Origin of Species", Darwin wrote that "When a species, owing to highly favourable circumstances, increases inordinately in numbers in a small tract, epidemics...often ensue". Disease and parasitism are two prevalent selective forces which influence natural rates of dispersion of species (Harper, 1977; Levin, 1978). If monoculture plantations of forest trees are to cope with such forces, the maintenance of high levels of genetic diversity will be essential (Kimmins, 1987).

Long-lived species, in particular, rely on genetic diversity to cope with changing environments (Allard and Bradshaw, 1964; Marshall, 1977; Adams, 1981; Namkoong, Kang and Brouard, 1988). The long generation time of white spruce renders it particularly vulnerable to such changes as global warming, introduction of foreign pests and diseases, forest fragmentation, and increasing pollution (Kimmins, 1987; Namkoong *et al*, 1988). Already, widespread susceptibility to introduced pathogens has imperiled numerous North American trees, including American chestnut

[*Castanea dentata* (Marsh.) Borkh] (Burnham, 1988), American elm (*Ulmus Americana* L.) (Gibbs, 1978), sugar pine (*Pinus lambertiana* Dougl.) (Ledig, 1992), eastern white pine (*Pinus strobus* L.) and western white pine (*Pinus monticola* Dougl. ex D. Don) (Mielke, 1943; Hoff, McDonald and Bingham, 1976), Port-Orford-cedar [*Chamaecyparis lawsoniana* (A. Murr) Parl.] (Shephard, 1975; Zobel, Roth and Hawk, 1985), and Florida Torreya (*Torreya taxifolia* Arn.) (McMahan, 1989). In addition, the even-aged stands promoted by artificial regeneration have proven more susceptible to pest infestations than their wild counterparts (Patton, 1961; Kinloch, 1972) Since many infestations threaten primarily young or establishing trees (Patton, 1961; Kinloch, 1972), young plantations are particularily vulnerable. Some forest tree species, such as lodgepole pine (*Pinus contorta* Dougl. ex. Loud) naturally regenerate in densely populated, even-aged stands (Hosie, 1979; Preston, 1989) and, as such, are likely adapted to these conditions. Species such as white spruce, however, tend to be recruited in the forest understory more sparsely, interspersed with other species, such as aspen (*Populus tremuloides* Michx.), paper birch (*Betula papyrifira* Marsh.) and balsam fir [*Abies balsamea* (L.) Mills] (Fowells, 1965; Stiell, 1976; Hosie, 1979; Cheliak *et al*, 1984), making it more difficult for pests specific to white spruce to be transmitted to saplings under natural conditions.

With an increasing human population, the demand for fibre-related products is likely to increase in the future. One means of protecting natural forests will be to more intensively manage plantations for increased growth yields. The high levels of genetic diversity present in current wild populations of conifers such as white spruce promise

potentially large gains in growth and stress-resistance characteristics through elite breeding programs (Carlisle and Teich, 1970; Nienstaedt and Teich, 1972). In conjunction with managed breeding programs, the maintenance of unmanaged ecological reserves of forest species will be equally vital to the preservation of future forest productivity (Kimmins, 1987). Should plantations become vulnerable to unforeseen stresses in the future, the wild ecotypes, possessing longterm fitness traits, will be available to supplement breeding programs (Kimmins, 1987).

In order to preserve potentially valuable genetic resources which currently exist in wild stocks of white spruce, it will be necessary to establish a means of efficiently assessing and monitoring genetic variation within both natural and managed populations. Microsatellite markers may be well suited for this purpose, because once identified, they are both cost effective and highly sensitive in detecting genetic differences between individuals, requiring only minute amounts of crude DNA from each specimen to be tested.

1.2.2 Evolutionary Distribution of Microsatellites

Microsatellites have been observed to be abundantly, and, in some cases, randomly dispersed throughout many higher eukaryotic genomes (Miesfeld, Krystal and Arnheim, 1981; Hamada, Petrino and Kakunaga, 1982; Jeang and Hayward, 1983; Tautz and Renz, 1984; Gross and Garrard, 1986; Smith and Devey, 1994; Wu and Tanksley, 1993; Hite, Eckert and Cheng, 1996), while being quite rare in organelle DNA, as well as in prokaryotes (Wang, Weber, Zhong and Tanksley, 1994). As DNA sequence information has become more available for a variety of genomes, it has become

apparent that microsatellites represent a particularly rich source of genetic polymorphism, present at high copy number within most higher organisms examined.

Microsatellites vary in both frequency and type between different species and groups of species. In mammals, the $(CA)_n$ repeat represents one of the most abundant microsatellites, occurring approximately once every 30 Kb (Weber and May, 1989), whereas in many plant species, $(CA)_n$ repeats appear to be among the least common of dinucleotide SSR's (Lagercrantz, Ellegren and Andersson, 1993; Wang *et al*, 1994; Bell and Ecker, 1993; Smith and Devey, 1994; Dow *et al*, 1995; Taramino and Tingey, 1996; Ma *et al*, 1996; Wu and Tanksley, 1993). The frequency of $(CT)_n$ repeats, on the other hand, was not found to vary significantly between plant groups and vertebrates. Though the $(CT)_n$ microsatellite comprises a small portion of vertebrate genome SSR's, it represents one of the most common plant microsatellites (Lagercrantz *et al*, 1993; Wang *et al*, 1994; Bell and Ecker, 1993; Smith and Devey, 1994; Dow *et al*, 1995; Taramino and Tingey, 1996; Ma *et al*, 1996; Wu and Tanksley, 1993). This reflects the overall lower frequency of microsatellites in plants as compared to vertebrate genomes studied to date, as well as the different composition of microsatellites present in different groups of organisms. Based on database studies of known sequence among divergent plant families, Lagercrantz *et al* (1993) found microsatellites to be five times less prevalent, overall, in plant species than in vertebrates.

The only dinucleotide repeat in plants found to significantly exceed the $(CT)_n$ repeat in frequency, perhaps on a par with the $(CA)_n$ repeat in

vertebrates, was the $(AT)_n$ repeat (Lagercrantz *et al*, 1993; Bell and Ecker, 1993). Unless discovered in database searches of sequence, however, this SSR has been impractical to screen for, due to $(AT)_n$ probes binding to themselves. In many cases, $(AT)_n$ SSR's have been found by chance, being the most common repeat to occur in compound plant microsatellites, next to other SSR's (Smith and Devey, 1994; Dow *et al*, 1995; Echt *et al*, 1996).

1.2.3 Assessment of Polymorphism at Microsatellite Loci

In 1989, several groups (Litt and Luty; Weber and May; Tautz) independently predicted that length polymorphisms among microsatellites might arise in a similar way to that previously reported for longer repeats of 11-60 nucleotides per unit, known as variable number tandem repeat (VNTR) loci (Nakamura, Leppert, O'Connell, Wolff, Holm, Culver, Martin, Fujimoto, Hoff, Kumlin and White, 1987), or minisatellites (Jeffreys, Wilson and Thein, 1985). In these highly polymorphic loci, such as were found in human RFLP markers by Wyman and White (1980), alleles were found to vary by the number of tandem repeats they contained. In order to assess whether microsatellites might also exhibit similar length polymorphisms, Litt and Luty (1989), Weber and May (1989), and Tautz (1989) used PCR to amplify specific microsatellite loci, the products of which were run on polyacrylamide gels, which would allow detection of even single nucleotide differences among samples.

All ten $(CA)_n$ human loci examined by Weber and May (1989) were found to exhibit length polymorphisms, differing by multiples of two bases, which corresponded to the repeat size. Similarly, Litt and Luty (1989) reported 12 allelic variants of a $(CA)_n$ repeat within the human cardiac muscle actin

gene, among 37 unrelated humans. Tautz (1989) assessed four microsatellite loci within three species, and reported finding four variants of a $(CAG)_n$ SSR within the Notch gene of *Drosophila*. In humans, among nine family members and three unrelated individuals, he found three alleles of a $(GT)_n$ repeat within the intergenic region between the human d and b globin genes, a locus earlier described by Miesfeld *et al* (1981). Within the same group of 12 humans, he also found six alleles of the intron region of the human cardiac actin gene. And among 10 pilot whales from a single population, Tautz observed three variants of a $(GA)_n$ repeat. Clearly, all three research groups found exceptionally high levels of polymorphism at each of the microsatellite loci examined.

In each of the above microsatellite studies, additional bands of reduced intensity were found to occur at multiple repeat intervals, above and below the most intense band. This *in vitro* gain or loss of repeats during PCR amplification has been found to be characteristic of microsatellite templates, and appears to illustrate *in vitro* a mutagenic process similar to that which occurs *in vivo*, albeit at greatly enhanced frequencies.

1.2.4 Classification of Microsatellites

Weber (1990) subdivided SSR's into three categories, based on his ability to predict their polymorphic values. He defined "perfect" microsatellites as those containing uninterrupted tandem repeats, such as $(CA)_{20}$, with no imperfections in repeat sequence, and lacking other adjacent repeats. "Imperfect" repeats consisted of tandem stretches of a repeat, containing at least one interruption of three or fewer consecutive, non-repeat bases. $(CA)_{10}$, AGA, $(CA)_8$ would be considered an imperfect repeat. Finally,

"compound" repeats were defined as two or more adjacent microsatellite repeats of at least 10 base pairs each, separated by three or fewer non-repeat bases. An example of a compound repeat is $(AT)_{12}$ TT $(CT)_{18}$.

Among 112 human microsatellites which Weber examined (1990), 64% were classified as perfect, 25% as imperfect, and 11% as compound. Most plant microsatellite studies have reported a much higher prevalence of compound repeats, most often with the microsatellite $(AT)_n$ accompanying another repeat. Among several tree species studied to date, for example, compound microsatellites have comprised 36.8% of 19 microsatellites identified in white pine (Echt *et al*, 1996), 46.7% of 15 microsatellites in bur oak (Dow *et al*, 1995), 50% of two loci examined in citrus species (Kijas et al, 1995), and 60% of five microsatellites in Monterey pine (Smith and Devey, 1994). Percentages of compound dinucleotide repeats in monocots varied from 44.5% of 76 microsatellites in wheat (Ma *et al*, 1996), and 23.1% of 13 microsatellites in rice (Wu and Tanksley, 1993), to a low of 8.8% among 34 loci in maize (Taramino and Tingey, 1996). 33% of six microsatellites in wild yam (Tarauchi and Konuma, 1994), and 23.5% of 17 loci in cultivated potato (Proven *et al*, 1996) were also compound. It is unclear why microsatellite composition should differ markedly between groups of organisms.

1.2.5 Informativeness of Microsatellites

Having subdivided SSR's into the above three categories, Weber found that the best predictor for the polymorphic information content (PIC) value of a given microsatellite was its longest stretch of uninterrupted, perfect repeats. Thus, the PIC value of compound SSR's could be predicted based on

the length of the longest continuous tract of perfect repeats, while even very lengthy imperfect repeats would be predicted to have relatively low PIC values.

In his studies, Weber (1990) found that human $(CA)_n$ loci with 10 or fewer repeats yielded no polymorphism, while PIC values of those with 11 to 15 repeats were quite variable. SSR's with 16 or more repeats consistently demonstrated moderate to high levels of polymorphism, increasing to a maximum of 21 repeats, after which PIC values leveled off.

Results in more recent plant studies have generally agreed with Weber's 1990 findings. However, some differences have emerged. In testing 32 pairs of primers on wheat microsatellites, Ma *et al* (1996) did not find a significant correlation between length of repeat and degree of polymorphism. Among the 22% of primers which produced single locus, polymorphic products, the range of polymorphic repeats, with n=13-18 for all loci examined except one, would agree generally with Weber's guidelines, though. The other polymorphic repeat, of n=8, would appear to be an exception to Weber's findings. Bell and Ecker (1993), likewise, failed to find a correlation between the length of *Arabidopsis* microsatellites and polymorphism. Some loci of n=13-16 exhibited four to six alleles among six strains tested, while other loci of n=23-25 produced only two or three alleles within the same strains. Wu and Tanksley (1993) did find the pattern of microsatellite polymorphism in rice to be similar to that described by Weber (1990), but with a slightly lower number of repeats (n=12) leading to consistent polymorphism.

In several plant species studied, informativeness was found to vary with the type of SSR being assessed. For microsatellites exceeding n=10 in maize, Taramino and Tingey (1996) found 13% of $(AC)_n$ repeats to be monomorphic, while all $(AG)_n$ repeats tested were polymorphic. Similarly, Bell and Ecker (1993) examined four types of SSR's in *Arabidopsis*. Of 37 $(GA)_n$ loci amplified, 83% were single locus and polymorphic. Among 22 $(CA)_n$ loci, in contrast, only 13.6% were found to be single locus, and of these, only one, or 4.5% of the total, was polymorphic. 40% of five $(AT)_n$ loci examined produced single locus, polymorphic alleles. And of four $(A)_n$ loci assessed, two produced single locus, monomorphic products, and two produced single locus, polymorphic products.

1.2.6 Microsatellite frequencies vary with repeat number

Microsatellite frequencies in a given genome may vary substantially, in accordance with the repeat number being screened for. Of 61 $(CT)_n$ SSR's of $n \geq 6$ located in the EMBL database of plant genomes by Lagercrantz *et al* (1993), only 8.2% exceeded 15 repeats, while 83.6% contained between six and 11 repeats. The $(CA)_n$ microsatellite in vertebrates appears to be an exception to the rule of declining frequencies of microsatellites with increasing repeat number, as Weber (1990) found little decrease in the frequency of $(CA)_n$ repeats for up to n=20. This has contributed to the feasibility of screening vertebrate genomes for microsatellite markers.

1.2.7 Distribution of Microsatellites within Plant Genomes

Plant microsatellites have also been found to differ from those characterized by Weber (1990) in the degree to which they amplify as single-locus markers. A significantly higher percentage of plant

microsatellites appear to be embedded within larger repetitive units than has been reported among vertebrate genomes. This results in complex banding patterns of PCR amplified microsatellites when examined on acrylamide gels. While such multi-locus markers may be useful for assessment of parentage, or for identification purposes, they must be scored as dominant markers, and, as such, are of limited value in population studies.

In his 1990 study, Weber found only 7% of human $(CA)_n$ microsatellites to amplify three or more alleles per locus. This contrasts with most plant studies to date. At the lower end of the scale, Provan *et al* (1996) found 13.6% of 22 primer pairs developed for cultivated potato to be multi-locus, and Wu and Tanksley (1993) reported 15.4% of the 13 loci tested on rice to be likewise. Ma *et al* (1996) reported "no distinct product" for 31% of 32 loci tested in wheat. In *Arabidopsis*, the $(CA)_n$ loci appeared to be mostly embedded within repetitive DNA, as 72.7% of 22 primer pairs produced complex banding patterns (Bell and Ecker, 1993). Approximately 25% of $(CT)_n$ microsatellites were reported to be multi-locus in several forest tree species studied to date, including bur oak (Dow *et al*, 1995), eastern white pine (Echt *et al*, 1996), and Monterey pine (Smith and Devey, 1994).

Schmidt and Heslop-Harrison (1996) examined the physical distribution of various microsatellites within *Beta vulgaris* L., a cultivar of sugar beet. *In situ* hybridizations of various SSR oligonucleotides on sugar beet chromosomes, as well as Southern hybridizations of the same oligonucleotides with in-gel, electrophoresed digests of sugar beet genomic DNA, allowed visual depictions of the distribution of these SSR's throughout

the sugar beet genome. Considerable variation in hybridization patterns was observed among different SSR's. Two dinucleotide repeats, $(GA)_n$ and $(TA)_n$, were found to be mostly evenly distributed throughout the genome, with minor amounts of repetitive bands evident on in-gel Southerns for both. In the case of $(GA)_n$, *in situ* hybridization depicted primarily even distribution throughout the chromosomes, with minor clustering at centromeres, while $(AT)_n$ was characterized by weak clustering throughout the chromosomes, and underrepresentation at centromeres. From the intensity of in-gel Southerns, $(GA)_n$ appeared to be the most abundant SSR analyzed, followed by $(CAC)_n$, $(GGAT)_n$, and $(TA)_n$, although it is likely that $(TA)_n$ signal was underrepresented due to the self-binding nature of this probe. $(CAC)_n$ and $(GGAT)_n$ SSR's were also quite evenly dispersed, with some banding on in-gel Southerns evident, as well as some localization of *in situ* signals away from centromeres for $(GGAT)_n$. Three other SSR's, $(GATA)_n$, $(CA)_n$ and $(GACA)_n$, hybridized primarily in banded patterns on in-gel Southerns, and, in the cases of $(CA)_n$ and $(GATA)_n$, localized strongly to centromeric regions *in situ*. The biased distribution of these SSR's may limit their usefulness as population genetic markers (Queller *et al*, 1993). This study reveals how varied the profiles of different SSR's can be within a single genome, and highlights the importance of characterizing individual microsatellite motifs with respect to representation and distribution among chromosomes for a species of interest, prior to embarking on large-scale attempts to isolate such markers for use in population studies. Ideally, population genetic markers should be both evenly dispersed throughout the genome (Queller *et al*, 1993), and present at sufficient density to allow inferences about the genome as a whole (Lewontin, 1974).

1.2.8 Mutational Mechanism of Microsatellites

The mutational mechanism widely believed to cause hypervariability at microsatellite loci involves slipped strand mispairing (Kornberg, Bertsch, Jackson and Khorana, 1964; Wells, Buchi, Kossel, Ohtsuka and Khorana, 1967*a*; Wells, Jacobs, Narang and Khorana, 1967*b*; Kornberg, 1980; Streisinger and Owen, 1985; Levinson and Gutman, 1987; Wolff, Plaetke, Jeffreys and White, 1989; Coggins and O'Prey, 1989; Strand, Prolla, Liskay and Petes, 1993; Hite *et al*, 1996). During DNA replication, DNA polymerase is believed to briefly dissociate and reassociate with its template, allowing an opportunity for reassociation with the wrong repeat unit, where strings of many identical tandem repeats occur (see Figure 1.3.b). Strand slippage appears to be dependent upon the length of the actual repeat unit (Wells *et al*, 1967b), as well as on the number of perfect tandem repeats present at a site (Streisinger and Owen, 1985; Levinson and Gutman, 1987b; Weber, 1990). Wells *et al* (1967b) observed that dinucleotide repeats incubated with DNA Polymerase I were much more likely to generate repeat expansion artifacts than tetranucleotide repeats, similarly incubated. Likewise, Levinson and Gutman (1987), and Blaisdell (1983) found runs of single base motifs, excess over that which would occur randomly in mammalian genome databases, at higher frequencies than excess runs of larger motifs.

Weber's (1990) observation that polymorphism of $(CA)_n$ loci in humans increases with the number of uninterrupted microsatellite repeats present agrees with *in vivo* studies in *Escherichia coli* in which dinucleotide expansion of $(A)_n$ repeats in T4 bacteriophage (Streisinger and Owen, 1995), and $(CA)_n$ repeats in m13 bacteriophage, (Levinson and Gutman,

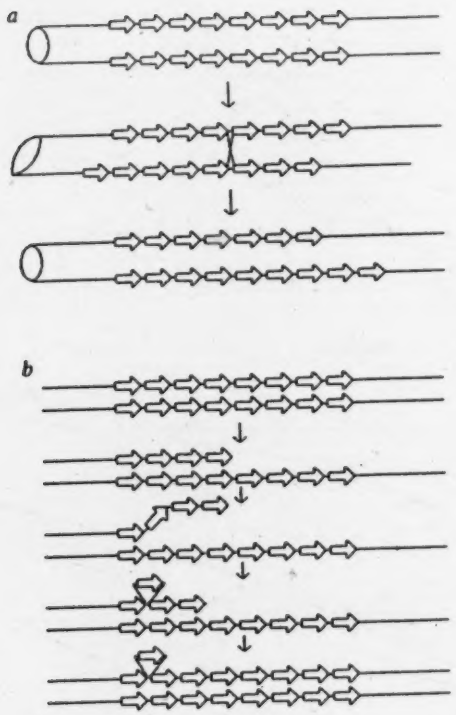

Figure 1.3 Mechanisms of mutation within tandem repeat sequences; (a) unequal cross-over; (b) slipped-strand mispairing (Strand, Prolla, Liskay and Petes, 1993)

1987b), was found to markedly increase for templates with 10 or more repeat units. In yeast, Strand *et al* (1993) noted that tract instability in meiosis was similar to that in mitosis, discounting the involvement of unequal crossover, and supporting strand slippage as a mutational mechanism. Earlier studies had also shown that mutation of genes involved in recombination had not affected simple repeat tract instability in *Escherichia coli* (Levinson and Gutman, 1985), or in yeast (Henderson and Petes, 1992).

While highly mutagenic in comparison to other loci, it is obvious that microsatellites must be much more stable *in vivo* than has often been witnessed *in vitro*. In an attempt to identify cellular factors which might contribute to enhanced stability, Strand *et al* (1993) examined several genes involved in both DNA mismatch repair and proofreading. They demonstrated that mutations in any of three genes involved in mismatch repair in yeast increased simple repeat tract instability 100-700 fold, whereas mutations in proof reading genes had little effect. This is consistent with the enhanced rate of mutagenesis of microsatellite loci found in PCR reactions, in which thermo-stabile polymerases may retain some level of proofreading function, but lack mismatch repair functions.

Mechanisms proposed to cause length polymorphisms in longer tandem repeats of 11-60 nucleotide units, as are found in minisatellites (Jeffreys *et al*, 1985) or VNTR's (Nakamura *et al*, 1987), or of 3-300 nucleotide repeat units found within malarial parasite antigen genes (Ravetch, Feder, Pavlovec and Blobel, 1984; Kemp, Coppel and Anders, 1987; Weber, 1988), have included unequal crossover during meiosis (Smith, 1976), slipped

strand mispairing during DNA replication (Streisinger and Owen, 1985; Overhauser, McMahan and Wasmuth, 1987; Wolff et al, 1989; Coggins and O'Prey, 1989; Luty, Willar, Ledbetter, Ledbetter and Litt, 1990; Strand *et al*, 1993), or some combination of the two (Levinson and Gutman, 1987*a*).

In PCR reactions involving microsatellites, an apparently accelerated rate of mutation leads to production of numerous "stutter" bands, differing in incremental steps of extra or fewer repeat units from the original template DNA. In an attempt to reduce these PCR generated artifacts, Hite *et al* (1996) experimented with thermophilic polymerases, including UITma, Pfu, Vent, and Deep Vent, as well as with various PCR conditions such as pH, dNTP concentrations, Mg++ concentration, and use of 3'Æ 5' exonucleases, with no significant reduction in artifact bands. Use of thermolabile polymerases, such as Sequenase 1.0, Sequenase 2.0, and 3'Æ 5' exonuclease deficient Klenow fragment at 37°C, however, dramatically reduced production of artifact bands. Hite *et al* (1996) speculated that it was the extension temperature, itself, during PCR which greatly accelerated the mutagenic process of slipped-strand mispairing, by destabilizing duplex tracts of microsatellite repeats, such that they had enhanced opportunities to re-align incorrectly during DNA synthesis. The much lower rates of mutagenesis observed at microsatellite loci in nature, in comparison to PCR, would be congruent with both the presence of DNA mismatch repair enzymes, and the lower temperatures at which *in vivo* DNA replication occurs, particularly in boreal forest species, such as white spruce. Hite *et al* (1996) discounted unequal cross-over as a mutational mechanism for generating microsatellite artifacts, because artifacts of intermediate length

did not spontaneously arise, but were rather always in steps corresponding to additions or losses of one repeat unit.

While highly mutagenic in comparison to other types of markers, microsatellites have been found to be stably inherited and to segregate in co-dominant, Mendelian fashion (Tautz, 1989; Weber and May, 1989; Bell and Ecker, 1993; Wu and Tanksley, 1993; Smith and Devey, 1994; Dow, et al, 1995; Kijas *et al*, 1995; Taramino and Tingey, 1996).

1.2.9 Distribution of Microsatellite Alleles

Weber and May (1989) noted that for most human $(CA)_n$ loci, alleles tended to cluster in steps of two nucleotide differences, within six nucleotides of the most predominant allele, and that in most cases, the entire range of alleles differed by no more than 20 nucleotides in length. Plant microsatellites have generally conformed to this pattern, with the exception that many have exhibited much wider ranges of alleles per locus. Slightly higher ranges of alleles per locus were found in grape, which averaged 24 nucleotides (Bowers *et al*, 1996), wheat with 28.5 nucleotides (Ma *et al*, 1996), maize with 38.2 nucleotides (Taramino and Tingey, 1996), and bur oak with 29.5 nucleotides (Dow *et al*, 1995). Some plants exhibited extremely large size differences among alleles, such as maize, for which alleles at many loci differed by 50 to 255 nucleotides in size (Taramino and Tingey, 1996). Other plant species conformed more closely to the range of alleles characterized by Weber for human microsatellites, including wild yam (Tarauchi and Konuma, 1994), Monterey pine (Smith and Devey, 1994), *Arabidopsis* (Bell and Ecker, 1994), and cultivated potato (Provan *et al*, 1996).

Trinucleotide repeats in plants generally exhibited larger size ranges in alleles than dinucleotide repeats. The average range of trinucleotide alleles in maize was 107.25 nucleotides per locus (Taramino and Tingey, 1996), and 59 nucleotides in wheat (Ma *et al*, 1996). A lower range of 21 nucleotides was observed in citrus trinucleotide alleles (Kijas *et al*, 1995).

1.2.10 Polymorphism Associated with Plant Microsatellites

Levels of polymorphism associated with microsatellites have in most cases greatly exceeded those found using other types of single locus markers. In only one study, that of Becker and Hewn, on barley (1995), did the authors not find microsatellite loci to be more polymorphic than RFLP's, with both averaging 2.1 alleles per locus, overall. Subsequent to this finding, however, Maroof, Biyashev, Yang, Zhang and Allard (1994), using four of Becker and Hewn's previous SSR primer sequences, claimed "extraordinary polymorphism" of wild and cultivated barley lines.

In general, gene diversity indices of polymorphic microsatellite loci have been at least double that observed for RFLP's or isozyme studies. Polymorphic SSR loci in Monterey pine, for example, averaged six alleles per locus among 40 trees analyzed, with an average observed heterozygosity (H_O) of 0.625 (Smith and Devey, 1994), in comparison to earlier isozyme findings of 1.8 alleles per locus, and an H_O of 0.08 (Moran, Bell and Eldridge, 1988). Similarly, among 16 white pine trees, four microsatellite loci averaged 5.4 alleles per locus, with an H_O of 0.515 (Echt *et al*, 1996), while isozyme studies of 18 loci among 300 white pine trees yielded only 1.96 alleles per locus, and an overall H_O of 0.176 (Beaulieu and

Simon, 1994). Most plant microsatellite studies to date have followed a similar pattern.

Even within this context, many plant microsatellite surveys have indicated exceptionally high levels of observed heterozygosity, in comparison to previous RFLP or isozyme studies. These include observed heterozygosities of 0.72 in bur oak (Dow *et al*, 1995); 0.86 in grape (Bowers *et al*, 1996); 0.79 in cultivated potato (Provan *et al*, 1996); 0.76 in maize (Taramino and Tingey, 1996); 0.84 (Morgante et al, 1994) and 0.87 (Rongwen et al, 1995) in soybean, versus a previous RFLP based H_O of 0.32 in soybean (Keim *et al*, 1992); and 0.79 in rice, versus earlier RFLP based estimates in rice which were seven to ten times lower (Wu and Tanksley, 1993). All of the SSR based H_O's examined exceeded 0.50. Indeed, at the lower end of the scale was wild yam, which, at 0.54, still exceeded a previous isozyme based H_O of 0.23 by more than a factor of two.

In some cases, gene diversity estimates may be exceptionally high due to authors including only those SSR's which are informative. For comparison purposes, however, this is likely balanced by a similar practice among those using alternate marker systems. Hamrick and Godt (1990) earlier noted that such a tendency caused estimates of genetic diversity within RFLP and isozyme studies to be artificially high.

1.2.11 Potential Uses of Microsatellites

The high levels of diversity at many microsatellite loci, along with their abundance, and neutrality make them well suited for population studies, as well as for mapping and identification purposes.

The ability to assess sequence tagged microsatellite loci efficiently is an asset in large-scale studies, necessary if baseline data for economically or ecologically important species, such as white spruce, are to be gathered. Previous white spruce studies have focused upon isozyme loci, which have seldom exceeded two or three alleles per locus. Forty seven white spruce trees surveyed at 14 isozyme loci, for example, led to mean heterozygosity of only 0.183 (Cheliak *et al*, 1984), in spite of this species being believed to be highly genetically diverse (Hamrick and Godt, 1990). This has made it difficult to detect subtle changes which may occur among populations as a result of management activities or environmental changes. The potentially greater sensitivity of microsatellite markers may facilitate monitoring of genetic resources, allowing detection of potential changes at an early stage.

These advantages, however, are contingent upon finding loci which are highly polymorphic, unique, and well dispersed throughout the genome. Since these characteristics can vary widely between species, it must be established separately for each species whether the pursuit of these markers is indeed worth the significant undertaking and cost required to develop them.

Introduction Literature Cited

Adams, W. T. 1981. Population genetics and gene conservation in pacific northwest conifers. pp. 401-415, *In* Evolution Today, Proceedings of the Second International Congress of Systematic and Evolutionary Biology, G. G. E. Scudder and J. L. Reveal (eds.), Hunt Inst. for Bot. Documentation, Carnegie-Mellon University, Pittsburg, PA.

Alden, J. and C. Loopstra. 1987. Genetic diversity and population structure of *Picea glauca* on an altitudinal gradient in interior Alaska. Can. J. For. Res. **17**: 1519-1526.

Anderson, E. 1949. Introgressive hybridization. Wiley, New York.

Barrett, J. A. 1981. The evolutionary consequences of mono-culture. pp. 209-248, *In* J. A. Bishop, and L. M. Cook (eds.), Genetic consequences of man made change. Academic Press, London.

Beaulieu, J. and J. P. Simon. 1994. Genetic structure and variability in *Pinus strobus* in Quebec. Can. J. For. Res. **24**: 1726-1733.

Becker, J. and M. Heun. 1995. Barley microsatellites: allele variation and mapping. Plant Molecular Biology, **27**: 835-845.

Beckman, J. S. and M. Soller. 1990. Toward a unified approach to the genetic mapping of eukaryotes based on sequence-tagged microsatellite sites. Bio/Technology, **8**: 930-932.

Bell, C. J. and J. R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics, **19**: 137-144.

Beineke, W. F. and T. O. Perry. 1966. Genetic variation in ability to withstand transplanting shock. In Proc. Eighth So. Conf. For. Tree Improv., Savannah, GA, pp. 106-109.

Blaisdell, B. E. 1983. A prevalent persistent global nonrandomness that distinguishes coding and non-coding eucaryotic nuclear DNA sequences. J. Mol. Evol., **19**: 122-133.

Bowers, J. E., G. S. Dangl, R. Vignani and C. P. Meredith. 1996. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome, **39**: 628-633.

Bradshaw, A. D. 1972. Some of the evolutionary consequences of being a plant. Evol. Biol. **5**: 25-47.

Brown, A. H. D. 1978. Isozymes, plant population genetic structure and genetic conservation. Theor. Appl. Genet. **52**: 145-157.

Brown, A. H. D. 1992. Human impact on plant gene pools and sampling for their conservation. Oikos, **63**: 109-118.

Burnham, C. R. 1988. The restoration of the American chestnut. Am. Sci. **76**: 478-487.

Callaham, R. Z. 1970. Geographic variation in forest trees, pp. 43-47. *In* O. H. Frankel, and E. Bennett (eds.), Genetic resources in plants - their exploration and conservation. F. A. Davis Co., Philadelphia.

Campbell, R. K. 1979. Genecology of Douglas-fir in a watershed in the Oregon Cascades. Ecology, **60**: 1036-1050.

Campbell, R. K. and F. C. Sorensen. 1984. Genetic implications of nursery practices. *In* Forest nursery manual: production of bareroot seedlings. M. L. Duryea, and T. D. Landis (eds), Nijhoff/Junk, The Hague, pp. 191-193.

Carlisle, A. 1977. The impact of man on the native pinewoods of Scotland. *In* Native pinewoods of Scotland, Bunce, R.G.H. and J.N.R. (eds.). Proc. Aviemore symp. Inst. Terrestrial Ecol., Cambridge, pp. 70-77.

Carlisle, A. and A. H. Teich. 1970. The costs and benefits of tree improvement programs. Can. For. Serv. Info. Report PS-X20. 28 p.

Cheliak, W. M., J. A. Pitel and G. Murray. 1985. Population structure and the mating system of white spruce. Can. J. For. Res. **15**: 301-308.

Cleary, B. C., R. D. Greaves and R. K. Hermann (eds.). 1978. *In* Regenerating Oregon's forests. Oregon State University Extension Service, Corvallis. 286 pp.

Coggins, L. W. and M. O'Prey. 1989. DNA tertiary structures formed *in vitro* by misaligned hybridization of multiple tandem repeat sequences. Nucleic Acids Res. **17**: 7417-7426.

Coles, J. F. and D. P. Fowler. Inbreeding in neighboring trees in two white spruce populations. Department of the Environment, Canadian Forestry Service, Maritimes Forest Research Centre, Fredericton, NB.

Conkle, M. T. 1973. Growth data for 29 years from the California elevational transect study of ponderosa pine. For. Sci. **19**: 31-39.

Day, P. R. 1973. Genetic variability of crops. Ann. Rev. Phytopathol. **11**: 293-312.

Dow, B. D., M. V. Ashley and H. F. Howe. 1995. Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. Theor. Appl. Genet. **91**: 137-141.

Drew, T. J. 1987. Excellence and results-oriented regeneration in Alberta's spruce forest: the place for cleaning and tending. pp 176-186, *In* F. McComb, chairman, Proc. Ind. Veg. Manage. Assoc. Alberta Semin. 18-19 March, 1987, Calgary, Alberta. Veg. Manage. Assoc. Alberta, Calgary, Alberta.

Echt, C. S., P. May-Marquardt, M. Hseih and R. Zahorchak. 1996. Characterization of microsatellite markers in eastern white pine. Genome, **39**: 1102-1108.

Edwards, A., Civitello, H., Hammond, H. A. and Caskey, C. T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. **49**: 746-756.

El-Kassaby, Y. A. 1992. Domestication and genetic diversity - should we be concerned? The Forestry Chronicle, Vol. 68, No.6.

Fowells, H. A. 1965. Silvics of forest trees of the United States. U. S. D. A. For. Serv. Agriculture Handbook No. 271, 762 p.

Frankel, O. H. 1974. Genetic conservation: Our evolutionary responsibility. Genetics, **78** (Suppl.): 53-65.

Frisque, G., G. F. Weetman and E. Clemmer. 1978. Reproduction and trial projected yields 10 years after cutting 36 pulpwood stands in eastern Canada. For. Eng. Res. Inst. Can., Pte. Claire, Quebec. Tech. Rep. TR-23.

Fryer, J. H. and F. T. Ledig. 1972. Microevolution of the photosynthetic temperature optimum in relation to the elevational complex gradient. Can. J. Bot. **50**: 1231-1235.

Gibbs, J. N. 1978. Intercontinental epidemiology of Dutch Elm disease. Ann. Rev. Phytopathol. **16**: 287-307.

Gross, D. S. and W. T. Garrard. 1986. The ubiquitous potential Z-forming sequence of eucaryotes, (dT-dG)*n* ◊ (dC-dA)*n* , is not detectable in the genomes of eubacteria. Mol. Cell. Biol. **6**: 3010-3013.

Hamada, H., Petrino, M. G. and Kakunaga, T. 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse genomes. Proc. Natl. Acad. Sci. U.S.A. **79**: 6465-6469.

Hamrick, J. L. and M. J. W.Godt. 1979. Levels of genetic variation in trees: influence of life history characteristics. *In* Proceedings of the symposium on Isozymes of North American Forest Trees and Forest Insects, Berkeley, CA; U.S.D.A. General Technical Report PSW-48.

Hamrick, J. L., Y. B. Linhart and J. B. Mitton. 1979. Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. Ann. Rev. Ecol. Syst. **10**: 173-200.

Harlan, J. R. 1975. Crops and man. American Society of Agronomy, Madison, Wisconsin.

Harper, J. L. 1977. Population biology of plants. Academic Press, London and New York.

Henderson, S. T. and T. D. Petes. 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Molec. Cell. Biol. **12**: 2749-2757.

Hite, J. M., K. A. Eckert and K. C. Cheng. 1996. Factors affecting fidelity of DNA synthesis during PCR amplification of $d(C\text{-}A)_n \Diamond d(G\text{-}T)_n$ microsatellite repeats. Nucleic Acids Res. **24**: 2429-2434.

Hoff, R. J., G. I. McDonald and R. T. Bingham. 1976. Mass selection for blister rust resistance: a method for natural regeneration of western white pine. U. S. For. Serv. Res. Note INT-202.

Hosie, R. C. 1979. Native Trees of Canada. 8th Ed. Minister of Supply and Services Canada; Fitzhenry and Whiteside Ltd, 150 Lesmill Road, Don Mills, Ontario, M3B 2T5.

Hughes, J. D., with collaboration of J. V. Thirgood. 1982. Deforestation, erosion, and forest management in ancient Greece and Rome. J. For. Hist. **26**: 60-75.

Jacob, H. J., K. Lindpainter, S. E. Lincoln, K. Kusumi, R. K. Bunker, Y. P. Mao, D. Ganten, V. J. Dzau and E. S. Lander. 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell, **67**: 213-224.

Jeang, K. T. and G. S. Hayward. 1983. A cytomegalovirus DNA sequence containing tracts of tandemly repeated CA dinucleotides hybridizes to highly repetitive dispersed elements in mammalian cell genomes. Mol. Cell. Biol. **3**: 1389-1402.

Jeffreys, A. J., V. Wilson and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature **314**: 67-73.

Jenkinson, J. L. 1984. Seed source lifting windows improve plantation establishment of Pacific Slope Douglas-fir. pp. 115-141, *In* Forest nursery manual: production of bareroot seedlings. M. L. Duryea and T. D. Landis (eds), Nijhoff/Junk, The Hague.

Keim, P., W. Beavis, J. Schupp and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. **85**: 205-212.

Kimmins, J. P. 1987. pp. 130-147, Forest Ecology. The University of British Columbia. Macmillan Publishing Company, New York.

Kuhnke, D. H. 1989. Silviculture statistics for Canada: an 11-year summary. For. Can., North. For. Cent., Edmonton, Alberta. Inf. Rep. NOR-X-301.

Kemp, D. J., R. L. Coppel and R. F. Anders. 1987. Repetitive proteins and genes of malaria. Annu. Rev. Microbiol. **41**: 181-208.

Kijas, J. M. H., J. C. S. Fowler and M. R. Thomas. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. Genome, **38**: 349-355.

King, J. N., B. P. Dancik and N. K. Dhir. 1984. Genetic structure and mating system of white spruce (*Picea glauca*) in a seed production area. Can. J. For. Res. **14**: 639-643.

Kinloch, B. B., Jr. 1972. Genetic variation in resistance to *Cronartium* and *Peridermium* rusts in hard pines. pp. 445-462, *In* Biology of rust resistance in forest trees: Proc. NATO-IUFRO Adv. Study Inst., Aug. 17-24, 1969, U.S.D.A. Misc. Publ. No. 1221, Washington, DC.

Kitzmiller, J. H. 1990. Managing genetic diversity in a tree improvement program. For. Ecol. Manage. **35**: 131-149.

Kleinschmidt, J. 1979. Limitations for restriction of the genetic variation. Silvae Genet. **28**: 61-67.

Kornberg, A. 1980. *In* DNA replication. W. H. Freeman, San Francisco.

Kornberg, A., L.L. Bertsch, J.F. Jackson and H.G. Khorana. 1964. Enzymatic synthesis of deoxyribonucleic acid. XVI. Oligonucleotides as templates and the mechanisms of their replication. Proc. Natl. Acad. Sci. USA, **51**: 315-323.

Kuhnke, D. H. 1989. Silviculture statistics for Canada: an 11-year summary. Information Report NOR-X-301, Forestry Canada, Northern Forest Centre, Edmonton.

Lagercrantz, U., Ellegren, H. and Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**: 1111-1115.

Ledig, F. T. 1992. Human impacts on genetic diversity in forest ecosystems. Oikos, **63**: 87-108.

Levin, D. A. 1978. The origin of isolating mechanisms in flowering plants. pp. 185-317, *In* Evolutionary biology, Vol. 11. M. K. Hecht, W. C. Steere and B. Wallace, (eds). Plenum Press, New York.

Levings, C. S., III. 1990. The Texas cytoplasm of maize: cytoplasmic male sterility and disease susceptibility. Science, **250**: 942-947.

Levinson, G. and G. Gutman. 1987a. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol Biol. Evol. **4**: 203-221.

Levinson, G. and G. Gutman. 1987b. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Res. **15**: 5223-5338.

Lewontin, R. C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.

Li, P., J. Beaulieu and J. Bousquet. 1993. Developing provisional seed transfer guidelines for white spruce in Quebec: a research outline. *In* Proceedings of the one-day symposium on research and development work on the genetics of white spruce in Quebec; Laurentian Forestry Centre; Information Report LAU-X-105B.

Litt, M. and Luty, J. A. 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Amer. J. Hum. Genet. **44**: 397-401.

Luty, J. A., Z. Guo, H. F. Willard, D. H. Ledbetter, S. Ledbetter and M. Litt. 1990. Five polymorphic VNTRs on the human X chromosome. Hum. Genet. **46**: 776-783.

Ma, Z. Q., M. Roder and M. E. Sorrells. 1996. Frequencies and sequence characteristics of di-, tri-, and tetra-nucleotide microsatellites in wheat. Genome, **39**: 123-130.

Mangelsdorf, P. C., R. S. MacNeish and W. C. Galinat. 1964. Domestication of corn. Science, **143**: 538-545.

Mangelsdorf, P. C. R. 1966. Genetic potentials for increasing yields of food crops and animals. Proc. Natl. Acad. Sci. U.S.A. **56**: 370-375.

Mangelsdorf, P. C. R. 1974. Corn. Its origin, evolution and improvement. Belknap Press, Harvard, Cambridge, Massachusets.

Maroof, S., R. M. Biyashev, G. P. Yang, Q. Zhang and R. W. Allard. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA, **91**: 5466-5470.

Marshall, D. R. 1977. The advantages and hazards of genetic homogeneity. Ann. New York Acad. Sci. **287**: 1-20.

Marshall, D. R. and A. H. D. Brown. 1975. Optimum sampling strategies in genetic conservation, pp. 53-80. *In* O. H. Frankel, and J. G. Hawkes (eds.), Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge.

McMahan, L. R. 1989. Conservationists join forces to save Florida Torreya. Center for Plant Conserv. **4**: 1-8.

Mielke, J. L. 1943. White pine blister rust in western North America. *In* Yale University Sch. For. Bull. No. 52. New Haven, CT.

Miesfeld, R., Krystal, M. and Arnheim, N. 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human delta and beta globin genes. Nucleic Acids Res. **9**: 5931-5947.

Millar, C. I. and W. J. Libby. 1989. Restoration: Disneyland or a native ecosystem? A question of genetics. Fremontia **17**: 3-10.

Moran, G. F. , J. C. Bell and K. G. Eldridge. 1988. The genetic structure and the conservation of the five natural populations of Pinus radiata. Can. J. For. Res. **18**: 506-514.

Morgante, M. and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**: 175-182.

Morgante, M., A. Rafalski, P. Biddle, S. Tingey and A. M. Olivieri. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome, **37**: 763-769.

Mulcahy, D. L. 1975. Differential mortality among cohorts in a population of *Acer saccharum* (Aceraceae) seedlings. Am. J. Bot. **62**: 422-426.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235**: 1616-1622.

Namkoong, G., H. C. Kang and J. S. Brouard. 1988. Tree Breeding: principles and strategies. Springer-Verlag, New York.

Nienstaedt, H. and A. Teich. 1972. The genetics of white spruce. U.S.D.A. For. Serv. Res. Pap. WO-15. 24 pp.

Overhauser, J., J. McMahan and J. J. Wasmuth. 1987. Identification of 28 DNA fragments that detect RFLPs in 13 distinct physical regions of the short arm of the chromosome 5. Nucleic Acids Res. **15**: 4617-4627.

Patton, R. F. 1961. The effect of age upon susceptibility of eastern white pine to infection by *Cronartium ribicola*. Phytopathology **51**: 429-434.

Preston, R. J. Jr. 1989. p. 37, North American Trees (4th Ed), Iowa State University Press.

Plucknett, D. L., N. J. H. Smith, J. T. Williams and N. Murthi-Anishetti. 1987. Gene banks and the world's food. Princeton University Press, Princeton, New Jersey.

Provan, J., W. Powell and R. Waugh. 1996. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor. Appl. Genet. **92**: 1078-1084.

Queller, D. C., J. E. Strassman and C. R. Hughes. 1993. Microsatellites and Kinship. Tree **8**: 285-289.

Rafalski, J. A. and S. V. Tingey. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends in Genetics, **9**: 275-280.

Ravetch, J. V., R. Feder, A. Pavlovec and G. Blobel. 1984. Primary structure and genomic organization of the histidine-rich protein of the malaria parasite *Plasmodium lophurae*. Nature (London), **312**: 616-620.

Rongwen, J., M. S. Akkaya, A. A. Bhagwat, U. Lavi and P. B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. **90**: 43-48.

Schmidt, T. and J. S. Heslop-Harrison. 1996. The physical and genomic organization of microsatellites in sugar beet. Proc. Natl. Acad. Sci. USA. **93**: 8761-8765.

Shepherd, C. J. 1975. *Phytophthora cinnamomi*, an ancient migrant to Australia. Search, **6**: 484-490.

Silen, R. R. and I. Doig. 1976. The care and handling of the forest gene pool. Pacific Search (June): 7-9.

Silen, R.R. 1978. Genetics of Douglas-fir. U.S.D.A. Forest Serv. Res. Pap. WO-35:1-34.

Silen, R. R. and J. G. Wheat. 1979. Progressive tree improvement program in coastal Douglas-fir. J. Forest. **77**: 78-83.

Smith, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science, **191**: 528-535.

Smith, J. S. C. 1988. Diversity of United States hybrid maize germplasm; isozyme and chromatographic evidence. Crop Sci. **28**: 63-69.

Smith, D. N. and M. E. Devey. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome, **37**: 977-983.

Stoehr, M. U. and Y. A. El-Kassaby. 1997. Levels of genetic diversity at different stages of the domestication cycle of interior spruce in British Columbia. Theor. Appl. Genet. **94**: 83-90.

Strand, M., T. A. Prolla, R. M. Liskay and T. D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature (London). **356**: 274-276.

Streisinger, G. and J. Owen. 1985. Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. Genetics. **109**: 633-659.

Taramino, G. and S. Tingey. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. Genome, **39**: 277-287.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res., **17**: 6463-6471.

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. **12**: 4127-4138.

Terauchi, R. and A. Konuma. 1994. Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. Genome, **37**: 794-801.

Theisen, P. A. 1980. Maintenance of genetic diversity and gene pool conservation of commercial or potentially commercial tree species in the Pacific Northwest Region United States Forest Service. pp.61-186, *In* Proc. servicewide workshop on gene resource menagement, Sacramento, CA, April 14-18, 1980. Timber Manage. Staff, U. S. For. Serv. Washington, DC, .

Thomas, M. R. and N. S. Scott. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor. Appl. Genet. **86**: 985-990.

Wang, Z., J. L. Weber, G. Zhong and S. D. Tanksley. 1994. Survey of plant short tandem DNA repeats. Theor. Appl. Genet. **88**: 1-6.

Weber, J. L. 1988. Molecular biology of malaria parasites. Exp. Parasitol. **66**: 143-170.

Weber, J. L. and P. E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44**: 388-396.

Weber, J. L. 1990. Informativeness of human (dC-dA)*n* /(dG-dT)*n* polymorphisms. Genomics. **7**: 524-530.

Wells, R. D., H. Buchi, H. Kossel, E. Ohtsuka and H. G. Khorana. 1967*a*. Studies on polynucleotides. LXX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeated tetranucleotide sequences. J. Mol. Biol., **27**: 265-272.

Wells, R. D., T.M. Jacobs, S.A. Narang and H. G. Khorana. 1967*b*. Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeat trinucleotide sequences. J. Mol. Biol., **27**: 237-263.

Wolff, R. K., R. Plaetke, A. J. Jeffreys and R. White. 1989. Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics. **5**: 382-384.

Wright, J. W. 1953. Pollen-dispersion studies: some practical applications. J. For. **51**: 114-118.

Wright, J. W. 1976. Introduction to forest genetics. Academic Press, New York.

Wu, K-S. and S. D. Tanksley. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. **241**: 225-235.

Wyman, A. R. and R. White. 1980. A highly polymorphic locus in human DNA. Proc. Acad. Sci. U.S.A. **77**: 6754-6758.

Zobel, D. B., L. F. Roth and G. M. Hawk. 1985. Ecology, pathology, and management of Port-Orford-cedar (*Chamaecyparis lawsoniana*). U. S. For. Serv. Gen. Tech. Rep. PNW-184.

Chapter 2 Materials and Methods

2.1 Population Sampling and Experimental Design

Two blocks, each consisting of adjacent planted, naturally regenerated, and old growth populations of white spruce, were sampled in the Prince Albert Model Forest, near Prince Albert, Saskatchewan. Within each of the six populations, needles were collected from at least thirty spruce trees at approximately 20 meter spacing, for a total of 180 trees sampled.

Core samples of old growth trunks placed their ages between 100 and 200 years. Records from Weyerhaeuser in Prince Albert indicated that the plantation was 18 years old at the time of sampling. The naturally regenerated trees were determined to be between 10 and 20 years of age.

2.2 DNA Extraction from White Spruce Population Samples

DNA was extracted from white spruce needles according to the CTAB protocol of Doyle and Doyle (1990). In this protocol, three grams of white spruce needles, washed in distilled water, rinsed in 95% ethanol, and dried, were frozen in liquid nitrogen, ground to a fine powder, and suspended in 12 ml of 2% CTAB solution (100 mM Tris HCl, pH 8.0, 1.4 M NaCl, 25 mM EDTA, 2% CTAB, 1% PVP40). After adding 50 ml of 40 mg/ml Proteinase K to the suspension, it was placed in a 60°C water bath for 30 minutes. An equal volume of chloroform was then added to each sample, mixed as gently as possible, and centrifuged at 2000 RPM for 15 minutes, after which the aqueous portion was poured into 2/3 volume of ice cold isopropanol,

causing the DNA to precipitate. In most cases, DNA could be lifted out with a glass hook. Otherwise, the suspension was gently spun, at approximately 200 rpm for five minutes, and the supernatent poured off. DNA was air dried on the bench top, then re-suspended in 400 ul of TE. The DNA was then treated with 2 ml of 20 mg/ml RNase A for 30 minutes at 37 ∞ C, chloroform extracted, dried and resuspended into approximately 400 ml of TE, depending on the size of the pellet.

2.3 Library Construction

2.3.1 DNA Extraction

For library construction, needles of several local white spruce trees were sampled from campus. DNA extraction was carried out as described above, with the exception that, after chloroform extraction from CTAB buffer, DNA was further purified on CsCl gradients, as described in Sambrook *et al*, 1989.

2.3.2 Preparation of Genomic Library Inserts

10 mg of genomic white spruce DNA was digested to completion with 15 units of Sau 3A in a volume of 50 ml, for three hours at 37∞ C. This digest was then run on a 2%, 1 x TAE minigel, alongside .65 mg of a 100 bp DNA ladder (Promega). White spruce DNA in the size range of 400 - 600 base pairs was cut out of the gel, and GENECLEANED (BIO 101) into a final volume of 70 ml, which was later diluted to 140 ml, to achieve most efficient ligation results.

2.3.3 Preparation of Vector for Ligation with Library Inserts

Five mg of pBluescript sk+ (Stratagene) was digested overnight with 20 U of BamH1 in a 25 ml volume. One ml of this digest was run on a 2%, 1 x TAE gel to test for completion of digest. To eliminate any traces of undigested vector prior to phosphatasing, the remaining 24 ml of digest was run on a 2%, 1 x TAE minigel, alongside both digested and undigested pBluescript. The lagging band, corresponding to the digested vector, was excised and gene-cleaned, in preparation for phosphatasing.

The BamH1 digested pBluescript SK+ vector (approximately 5 pmoles) was then incubated with 0.5 U of Calf Intestinal Phosphatase (0.5 unit being sufficient to phosphatase 50 pmoles of substrate), in a 50 ml volume at 37∞ C for 30 minutes. Afterwards, the Calf Intestinal Phosphatase was deactivated by heating to 75∞ C for 15 minutes, and then microfuged through a Millipore Ultrafree-Probind 0.45 mm filter unit to physically remove all previous enzymes prior to ligating.

2.3.4 Ligation of White Spruce Genomic Inserts into pBluescript

For each preparation of library insert and phosphatased vector, ligations were empirically tested using vector: insert molar ratio's of approximately 1:1, 1:2, and 1:3. In most cases, a vector: insert molar ratio of 1:2 was found to yield the highest percentage of positive transformants.

Ligation reactions were carried out in 10 ml volumes, which included 2U of T4 Ligase, 1 ml of 10 mM rATP, approximately 0.2 pmole of insert, and 0.1 pmole of Bam digested, phosphatased vector, at 4∞ C, overnight.

2.3.5 Preparation of $CaCl_2$ Competent Cells

This protocol was modified from that described in Sambrook *et al* (1989). *Escherichia coli* XL1 Blue MRF' (Stratagene) was plated on LB containing 50 µg/ml tetracycline, to ensure retention of the F' episome, and grown overnight at 37°C. Single colonies of 2-3 mm in diameter were transferred into 100 ml of LB broth + 50 µg/ml tetracycline, and incubated for approximately three hours at 37°C with vigorous shaking (300 cycles/min). The OD_{600} was monitored every 20-30 minutes, to ensure that cell density did not exceed 10^8 cells/ml. When cell growth had reached early log phase(OD_{600} = approximately .600), the broth was aseptically transferred to sterile, disposable 50 ml tubes, chilled to 0°C by placing on ice for 10 minutes, and centrifuged at 4000 rpm for 10 minutes at 4°C in a Sorvall GS3 centrifuge. After decanting media, pellets were resuspended in 10 ml of ice cold 0.1M $CaCl_2$, and centrifuged again at 4000 rpm for 10 minutes at 4°C. Media were again decanted, and pellets resuspended in 2 ml of ice-cold 0.1M $CaCl_2$ per 50 ml of original culture. 50% glycerol was added to 15% final concentration, and competent cells stored in 0.5 ml aliquots at -70°C for later use.

2.3.6 Transformation of Ligated Inserts into XL1-Blue.MRF'

Ligation products were transformed into XL1-Blue.MRF' according to the $CaCl_2$ method, as described by Sambrook *et al* (1989). 100 µl of competent cells were added to each 10 µl ligation reaction, mixed gently, and stored on ice for 30 minutes. Reactions were then transferred to a 42°C water bath for 60 seconds, followed by rapid chilling on ice for one to two minutes. Pre-warmed LB was added to a final volume of one ml per tube, then incubated for 45 minutes at 37°C, with vigourous shaking (up to 225

cycles/minute), to allow bacteria to recover and express ampicillin resistance. Upon completion, transformed reactions were microfuged at 1000 rpm for 10 minutes, and the volume of each reduced from 1000 µl to 200 µl. Several alliquots of the condensed suspensions were plated on 100 mm agar plates containing LB plus ampicillin, and grown overnight at 37°C to determine titres. During this time, the unplated transformants were stored at 4°C. After determining titres, remaining transformants were plated at the relatively low density of approximately 1000 colonies per 150 mm, in order to enhance signal of positive clones, and to facilitate isolation of positive colonies for second screenings.

To select against untransformed host cells, LB plating medium included 100 µg/ml ampicillin. 40 µl of 20 mg/ml X-Gal and 4 µl of 200 mg/ml IPTG were applied to the surface of each plate prior to plating transformants, to check for the presence of inserts in transformed cells. In most cases, there were fewer than 30% blue colonies on library plates. Since the inserts in this library were small, it can be expected that approximately 30% of them would allow in-phase reading of the B-Galactosidase genes, so that, in fact, most of the blue colonies would also be expected to contain library inserts.

After plating, colonies were grown at 37° C for approximately 16 hours, or until colonies were approximately 1 mm in diameter. Plates were then refrigerated at 4° C overnight, in preparation for colony lifts.

2.4 Screening of Genomic Library for Microsatellite Repeats

2.4.1 Colony Lifts of Genomic Library

Colony lifts were performed as described in *Current Protocols in Molecular Biology* (1989). Nylon reinforced nitrocellulose membranes were pressed firmly and evenly onto the surface of library plates, punctured near the perimeter in three places with a sterile needle, for later re-orientation of membranes to library plates, and left for five minutes on the bench top. Membranes were then lifted and placed, colony side up, onto a 3 MM Whatman sheet saturated with 0.5 M NaOH for seven minutes. Next, they were lifted and placed similarly onto a 3 MM Whatman sheet saturated with 1 M Tris Chloride, pH 7.6, for five minutes, followed by five minutes on a sheet saturated with 1 M Tris Chloride, pH 7.6, and 1.5 M NaCl. This was followed by two five minute rinses in the same solution, to remove bacterial debris.

Membranes were then placed onto dry 3MM Whatman paper and baked at 80 ° C for two hours.

Two sets of colony lifts were performed for each library plate. Prior to the second lift, each library plate was returned to the 37°C incubator for two hours to allow colonies to regenerate.

2.4.2 End-Labeling of Oligonucleotide Probes

65 picamoles of each of three oligonucleotides, $(CT)_{12}$, $(TAT)_8$, and $(TCT)_8$, were incubated with 66 picamoles, or 200 µCi, of [γ^{32}P] , and 50 units of T4

Kinase in a volume of 162 µl, at 37 ° C for 30 minutes (*Current protocols in molecular biology*, 1989).

After incubation, unincorporated [γ^{32}P] was removed by filtering the reaction through a Sephadex G-25 (Pharmacia NAP-5) column, from which labeled oligonucleotides were selectively eluted into one ml of 10 mM sodium phosphate buffer, pH 6.8. The entire ml of labeled probe was included in 100 ml of hybridization solution.

2.4.3 Prehybridization of Colony Lifts

Prehybridization was performed as described by Paetkau and Strobeck (1994). After baking at 80° C, membranes were placed in a solution of 2 x SSC, .1% SDS, 50 µg/ml Proteinase K (10 ml per membrane), and agitated gently at 37° C for two hours to overnight in a sealed plastic container.

Membranes were then rinsed in 2 x SSC at room temperature for five minutes twice, and either dried on 3MM Whatman paper for later hybridization, or placed directly into the hybridization solution.

2.4.4 Hybridization of Colony Lifts to Oligonucleotide Probes

Hybridizations were carried out as described by Paetkau and Strobeck (1994). Membranes were hybridized overnight at 50°C in a solution of 4% BLOTTO (5% skim milk powder, 0.02% Na azide), 6 x SSC, 0.05% Na pyrophosphate, and labeled probe, using enough solution to ensure that all membranes were fully immersed (approximately 10 ml per membrane, for 137 mm membranes). The following day, membranes were washed at room temperature twice in 2 x SSC, 0.05% Na pyrophosphate for 5 minutes each,

followed by a third wash in the same solution at 62° C for 30 minutes. In some cases, a fourth wash at 62° C was added, if the radioactive signal remained high.

Once radioactive signal was reduced sufficiently, membranes were placed on 3MM Whatman paper, wrapped in saran wrap, and placed on film in a cassette for overnight to two days, depending on strength of signal.

2.4.5 Identification of Positive Colonies from Genomic Library

The puncture marks on membranes were used to mark and align exposed film from the two sets of membranes. Those colonies which appeared clearly positive on both sets of membranes were re-plated at a lower density on fresh plates, for further screening. To accomplish this, individual colonies were suspended in one ml of LB + 100 µg/ml Ampicillin, and several aliquots, from 5 µl to 30 µl, plated on 1/4 150 mm plates of LB + Ampicillin, X-Gal, and IPTG. Colony lifts were repeated on plates which had appropriate density of single colonies to allow unambiguous identification of positives (preferably 25-60 colonies per 1/4 150 mm plate) (Ausubel *et al*, 1989). Colonies which appeared unambiguously positive in the second screening were prepared for commercial sequencing.

2.5 Preparation of Template for Commercial Cycle Sequencing

Colonies identified as positive in the second screening were resuspended in 400 µl of sterile ddH_2O, boiled for five minutes, and 2 µl of this used as template in a 25 µl PCR reaction, using T3 and T7 primers, which act as forward and reverse primers to the insertion site in pBluescript. Each PCR

reaction included 25 ng of each primer, 200 µM of each dNTP, 1.0 mM $MgCl_2$, 1 x Taq buffer (Perkin-Elmer), and 0.5U Taq polymerase. The amplification program, on a Perkin-Elmer GeneAmp thermo-cycler, was as follows:

94° C / 2'

(96° C / 30", 55° C / 1', 72° C / 30") 35 x

72° C / 5'

4° C / hold

Four 25 µl PCR reactions were completed for each sequencing template, pooled, and GENECLEANED (BIO 101) into a final volume of 15 µl. After checking concentrations of GENECLEANED (BIO 101) templates on a 2% agarose gel alongside known concentrations of 100 bp ladder (Promega), approximately four µl of this concentrated template was used as template in each sequencing reaction.

2.6 Design of Primers flanking Microsatellite Sequences

For those clones containing sufficient flanking sequence next to microsatellites of 11 to 30 repeats, primers were designed using OLIGO 4.1 Primer Analysis Software (National Biosciences, Inc.). Primer length and G/C content was adjusted to result in melting temperatures for both primers in a given pair being within one degree Celcius of each other, and for annealing temperatures to be as high as possible, exceeding 50° C in all cases.

2.7 PCR Assessment of Microsatellite Primers on White Spruce Population Samples

One primer of each pair was end-labeled with [γ^{32}P]dATP prior to PCR, using the following protocol:

40 µl 25 ng/µl primer (160 pmol)

4 µl [γ-32P] dATP, 3000 mM/Ci

1 µl T4-Polynucleotide Kinase (10 U/µl)

5 µl 10 x Kinase buffer

---

50 µl total volume

This reaction was incubated at 37° C for one hour, then heat deactivated at 75° C for 10 minutes.

Prior to assessing microsatellite loci, PCR conditions were optimized for each locus. This was done by varying concentrations of $MgCl_2$, in 0.5 molar increments, from one to three millimolar, as well as by varying annealing temperatures from 50-58° C, and extension times at 72° C from 1-30 seconds. The initial denaturing temperature of 94° C for two minutes was not varied.

The following PCR program was found to best amplify microsatellite sequences from all loci assessed, when using genomic template:

94° C / 2'

(96° C/ 20", 55° C / 30", 72° C / 1") 35 x

72° C / 5'

4° C/ hold

Hot starts were used in all cases. The 25 µl PCR cocktail included 25 ng of

template, 25 ng of each primer, 200 µM dNTP's, 2.0 mM $MgCl_2$, 0.5 U Taq polymerase, 50 mM KCl, 10 mM Tris (pH 8.0), and 0.1% gelatin.

1-3 µl of each PCR reaction was run on a 6% polyacrylamide/8M Urea/1xTBE gel at 75 watts for approximately three hours. ^{35}S-radio-labelled sequencing reactions of M13 template were run alongside microsatellite lanes, as size markers. Gels were dried under vacuum at 80° C for two hours before being placed on film for overnight to two days, depending on strength of signal. Loci exhibiting three or more alleles per individual were not pursued further for population analysis.

2.8 Estimate of Quantity of White Spruce Genome Screened

Approximately 16,000 library colonies were screened for $(CT)_n$ microsatellites. To be conservative, I estimated that 20% of these did not contain white spruce inserts, based on approximately 30% of library colonies being blue, which implies either no interruption of the B-galactosidase genes within pBluescript, or interruption by a relatively short insert which does not disrupt in-frame reading of these genes. PCR-based size estimates of library inserts, after subtracting the 150 bp of plasmid sequence which lies between T3 and T7 primers, averaged 550 bp. Using these estimates, approximately (16,000 colonies x .80 transformation efficiency x .550 Kb genomic DNA/clone), or 7040 Kb of the white spruce genome was screened for microsatellite repeats.

Materials and Methods Literature Cited

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl (Eds.). 1989. *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley Interscience, New York.

Doyle, J. J., and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus, **12**: 13-15.

Morgante, M., A. Rafalski, P. Biddle, S. Tingey, and A. M. Olivieri. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome, **37**: 763-769.

Paetkau, D., and C. Strobeck. 1994. Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology, **3**: 489-495.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Chapter 3 Results

3.1 Frequency and Composition of $(CT)_n$ Microsatellites in White Spruce

Within approximately 7040 Kb of white spruce genome screened, 10 microsatellites having a minimum of 11 consecutive repeats were identified, averaging one microsatellite per 704 Kb. Three of these microsatellites were too extensive in length (exceeding 60 consecutive repeats), and too embedded within repetitive DNA to obtain suitable flanking sequence for primer construction. Primers were designed for the remaining seven microsatellites, which averaged 18 perfect, consecutive repeats in length (Tables 3.1 and 3.2). Altogether, six of the ten microsatellites sequenced appeared to be associated with repetitive DNA. Five microsatellites were also compound, with $(AT)_n$ repeats comprising the most common adjacent repeats.

3.2 Informativeness of White Spruce Microsatellites

Microsatellite loci in this study were identified according to the library clones from which they were isolated. Of the seven primer pairs tested, one pair, HC 11 and 12, failed to effectively amplify the target sequence, while yielding large numbers of artifact bands across the length of the gel. Primers flanking this repeat were re-designed, with no noticeable difference in results. This locus, L4.1.b, was not pursued further. A second locus, L3.4.b, amplified by primers HC 1 and 2, yielded very polymorphic PCR products in the appropriate size range, along with multiple other polymorphic products, indicative of multi-locus amplification (Figure 3.1).

Table 3.1 White Spruce Microsatellites

Microsatellite repeats	Primer pairs	Expected PCR product length	Locus (Library clone)
$(CT)_{17} + (TA)_{14} + (T)_{13}$	HC1, HC2	260 bp	L3.4.b
$(CT)_{15}$	HC3, HC4	210 bp	L7.2.a
$(CT)_{11}$	HC5, HC6	143 bp	L8.2.b
$(CT)_{20}$	HC7, HC8	180 bp	L15.3.a
$(CT)_{17}$	HC9, HC10 HCO9, HCO10	230 bp 160 bp	L3.1.a L3.1.a
$(CT)_{22}$	HC11, HC12 HCO11, HCO12	235 bp 210 bp	L4.1.b L4.1.b
$(CT)_{23} + (GT)_{15}$	HCO13, HCO14	200 bp	L8.1.b
$(CT)_{60+}$	*n/a	n/a	L1.1.b
$(CT)_{60+}$	*n/a	n/a	L14.1.a
$(CT)_{60+}$	*n/a	n/a	L15.1.b

* not applicable, refers to microsatellites for which primers were not developed

Table 3.2 White Spruce Microsatellite Primers

Primers	Primer sequences	Optimal anneal temp.(°C)*	T_M (°C)*
HC 1	GATATAGTTAGCAGACGAACCAATGGA	50.2	56.3
HC 2	TAAGGCGGAGGTTCTGGGGTCA		61.5
HC 3	TTGAAAAAGAGGTTAGGAAGGGA	51.4	53.2
HC 4	TTCTTAAAGAAGCAGGGCATTG		53.2
HC 5	GTTAGGGTTTTGGACCTTTAGGGT	50.0	55.1
HC 6	ATCTGGCTAGAATGGTGGTAGCTT		54.5
HC 7	GACCTCTCCCAATAATGCGCCTC	54.5	59.8
HC 8	GCCCTGCCAGTGGAAGAAACA		58.7
HCO 9	AGAAAGTCCTCTCCCAAAAATG	44.3	51.1
HCO 10	ACTTCCACACGATCCATAATTG		50.9
HC 9	GTCCTCTCCCAAAAATGCGCCTC	55.4	61.8
HC 10	TCCCCTTTTGCGGATTGGTCA		61.0
HCO 11	CATTATCCCCACCATTGAACTC	46.3	52.6
HCO 12	GGAGGAAACACTAACCCTATACTTCA		53.5
HC 11	TCTTGGTAAATACATTATCCCCACCA	50.1	56.8
HC 12	CCATGGGAGGAAACACTAACCCT		57.1
HC 13	ACAAATACAGAAAGTCCTCTCCCA	51.0	53.7
HC 14	AACCCTGCCAGTGAAAAGACA		53.0

*Optimal annealing temp.'s and melting temp.'s were determined using the software program **Oligo** (National Biosciences, Inc.).

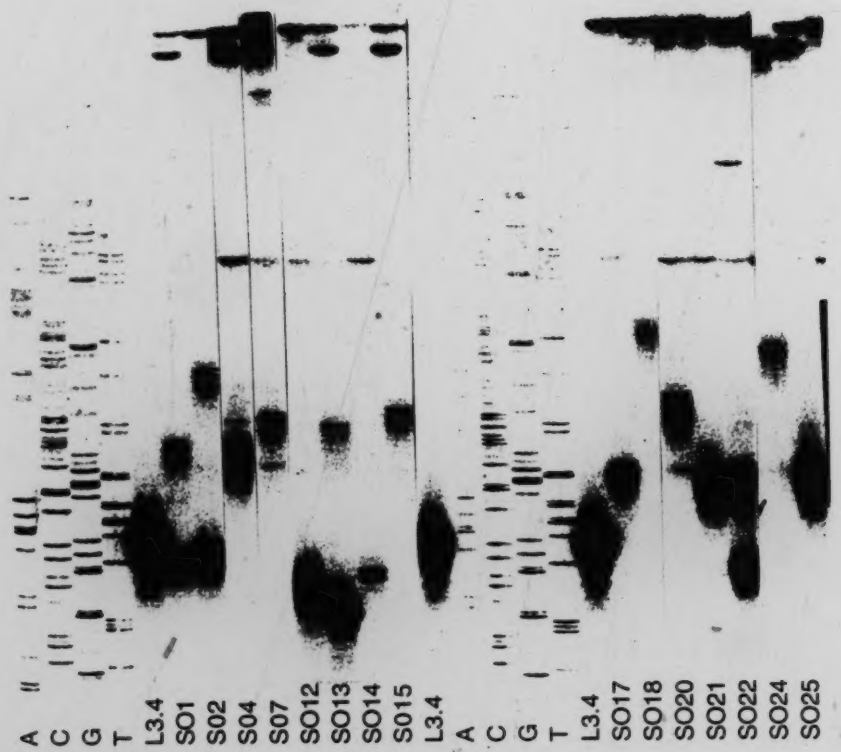

Figure 3.1. PCR amplification products generated by primers HC 1&2 upon locus L3.4.b, within individual tree samples from the **Snowfield Road Old Growth** population, and resolved on a denaturing polyacrylamide gel. Sequencing reactions of M13 were used as size standards.

This locus may be useful for identification purposes or paternity analysis, but was not pursued further as a population marker in this study.

Locus L7.2.a also generated polymorphic products in the expected size range. Because banding was localized to one region of the polyacrylamide gel, further attempts were made to optimize PCR conditions for this locus, in order that population samples might be assessed. Complexity of banding patterns, along with presence of extensive stutter bands, however, precluded definitive assignment of individual alleles (see Figures 3.2 - 3.7). It is likely that this locus is also repetitive in the genome. Attempts to interpret this locus were further complicated by the overriding presence of one genotype in all populations. Considering the diversity of alleles present, the strong presence of this apparently heterozygous genotype in the absence of homozygotes of either predominant allele was curious.

The four remaining primer pairs amplified products in the expected size range, consistent with single locus patterns. Amplification of 20 old-growth population samples, however, yielded only monomorphic products. Attempts to optimize PCR conditions by altering $MgCl_2$ concentration, annealing temperatures and times, or extension times failed to reveal any polymorphism within these loci, L3.1.a, L8.1.b, L8.2.b, and L15.3.a. A second pair of primers was designed to flank the $(CT)_{17}$ repeat within locus L3.1.a. It also amplified only monomorphic products.

Characteristics of the ten microsatellites identified in white spruce are summarized in Table 3.3.

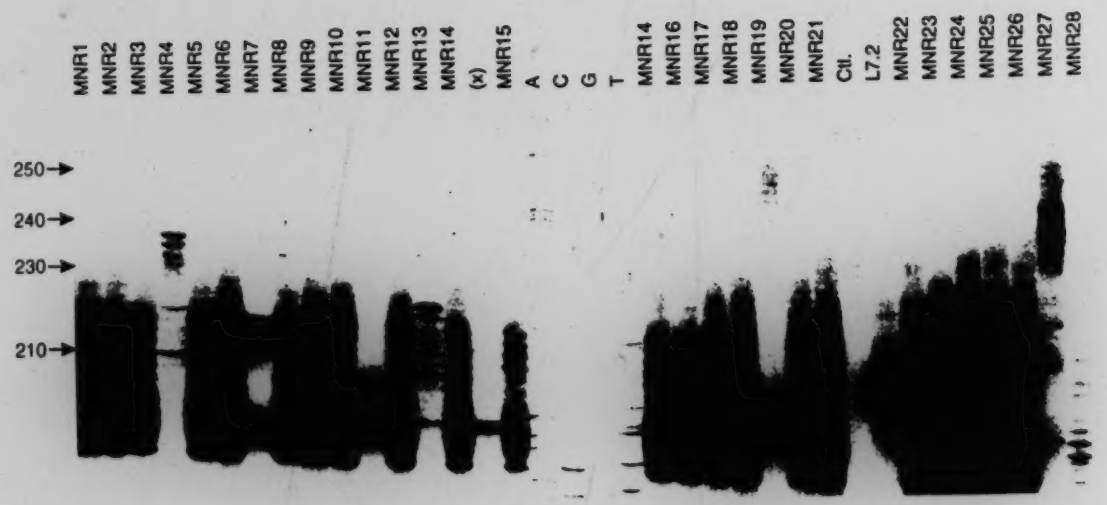

Figure 3.2. PCR amplification products generated by primers HC 3&4 on 28 individual tree samples from the **Montreal Lake Natural Regeneration** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

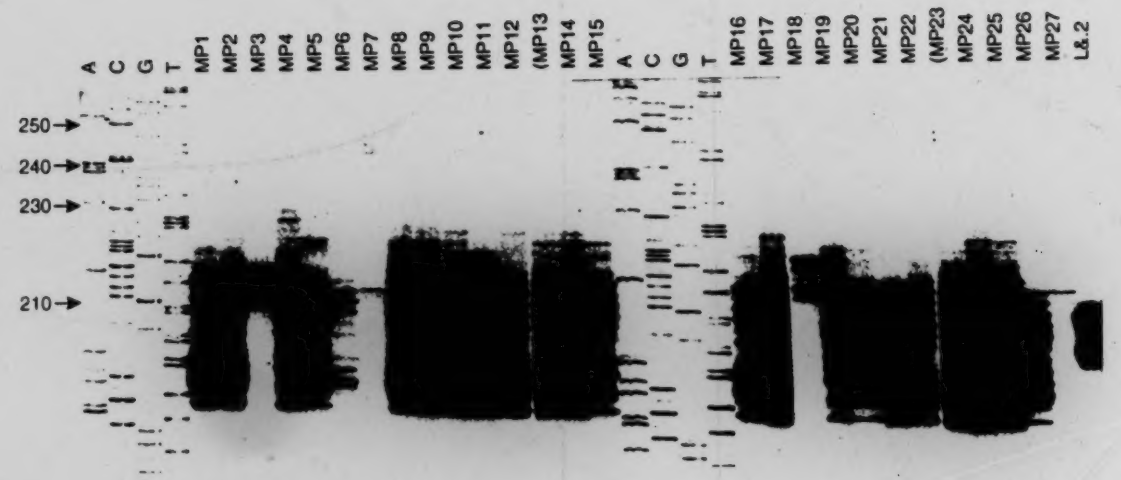

Figure 3.3. PCR amplification products generated by primers HC 3&4 on 27 individual tree samples from the **Montreal Lake Planted** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

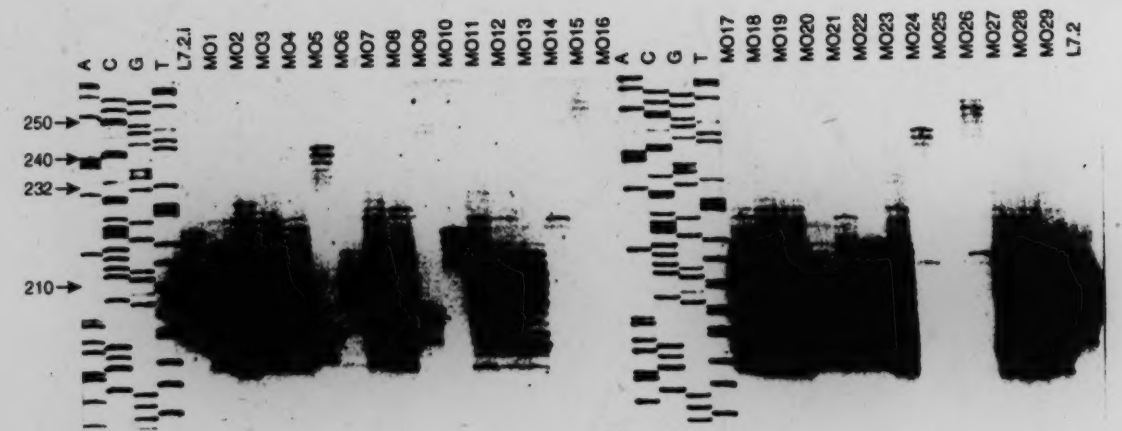

Figure 3.4. PCR amplification products generated by primers HC 3&4 on 29 individual tree samples from the **Montreal Lake Old Growth** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

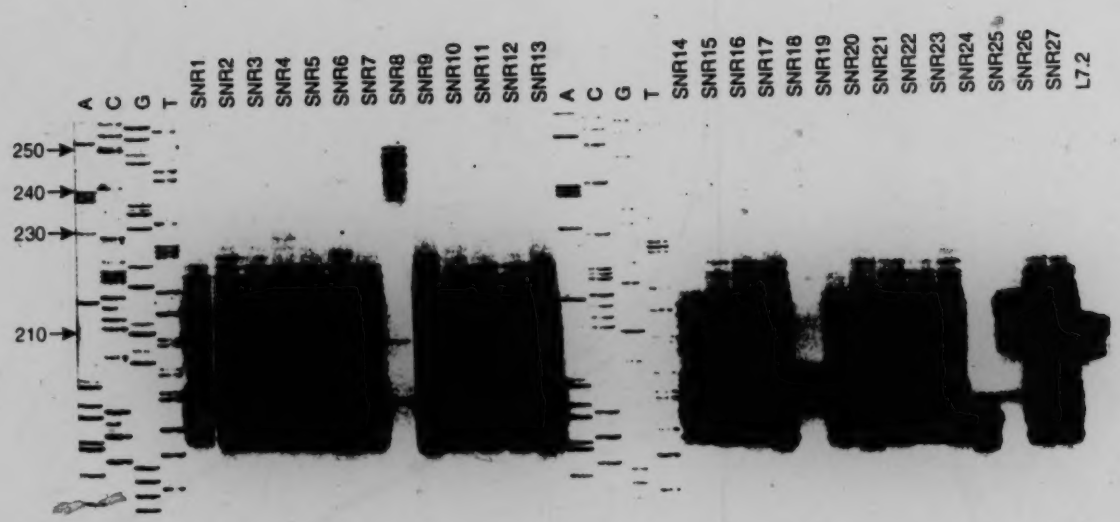

Figure 3.5. PCR amplification products generated by primers HC 3&4 on 27 individual tree samples from the **Snowfield Road Natural Regeneration** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

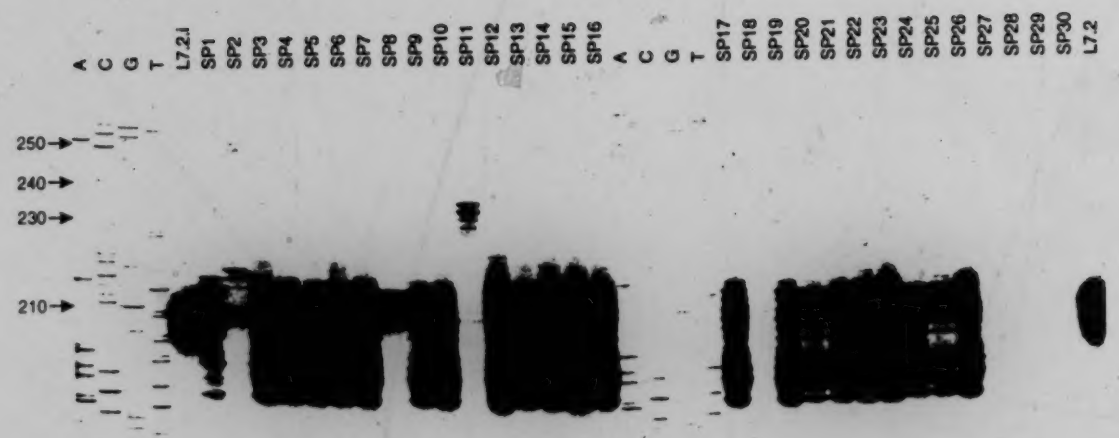

Figure 3.6. PCR amplification products generated by primers HC 3&4 on 30 individual tree samples from the **Snowfield Road Planted** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

Figure 3.7. PCR amplification products generated by primers HC 3&4 on 28 individual tree samples from the **Snowfield Road Old Growth** population, and resolved on a denaturing polyacrylamide gel. L7.2 refers to the library clone from which this locus was isolated. Sequencing reactions of M13 were used as size standards.

Table 3.3 White Spruce Microsatellites

Microsatellite repeats	Primer pairs clone)	Locus (Library	Amplification pattern
$(CT)_{17} + (TA)_{14} + (T)_{13}$	HC1, HC2	L3.4.b	multi-locus, polymorphic
$(CT)_{15}$	HC3, HC4	L7.2.a	multi-locus, polymorphic
$(CT)_{11}$	HC5, HC6	L8.2.b	single locus, monomorphic
$(CT)_{20}$	HC7, HC8	L15.3.a	single locus, monomorphic
$(CT)_{17}$	HC9, HC10	L3.1.a	single locus, monomorphic
	HCO9, HCO10	L3.1.a	single locus, monomorphic
$(CT)_{22}$	HC11, HC12	L4.1.b	multi-locus, unknown
	HCO11, HCO12	L4.1.b	multi-locus, unknown
$(CT)_{23} + (GT)_{15}$	HCO13, HCO14	L8.1.b	single locus, monomorphic
$(CT)_{60+}$	*n/a	L1.1.b	n/a
$(CT)_{60+}$	*n/a	L14.1.a	n/a
$(CT)_{60+}$	*n/a	L15.1.b	n/a

* not applicable, refers to microsatellites for which primers were not developed

Chapter 4 Discussion

4.1 Characterization of Microsatellites in White Spruce

Microsatellites in white spruce have differed from those characterized in other species in several ways. Weber (1990) reported an average of one microsatellite of at least 10 repeat units in length per 30 Kb of human genome. Lagercrantz *et al* (1993) later found microsatellites to be five times less abundant, overall, in plant species than in animal species. At a frequency of approximately one $(CT)_n$ microsatellite of at least 11 repeats per 704 Kb, white spruce microsatellites occurred much less frequently than have been reported in other plant species. Only *Pinus radiata*, estimated to have one dinucleotide repeat per 750 kb of DNA, has been described as having a similar frequency (Smith and Devey, 1994).

Informativeness of microsatellites in white spruce also contradicted patterns identified in other species. Weber (1990) reported length of uninterrupted microsatellite repeats in humans to be highly predictive of informativeness, with repeats of 16 or more units being consistently polymorphic. This finding has been upheld in many other species, including black bear (Paetkau and Strobeck, 1994), rice (Wu and Tanksley, 1993), wild yam (Terauchi and Konuma, 1994), *Citrus* (Kijas *et al*, 1994), bur oak (Dow *et al*, 1995), Monterey pine (Smith and Devey, 1994) and maize (Taramino and Tingey, 1996). In some species, such as soybean (Morgante, Rafalski, Biddle *et al*, 1994) and grape (Bowers *et al*, 1996), much shorter microsatellite repeats have also been highly polymorphic.

In white spruce, no correlation between number of consecutive repeat units and informativeness was evident. Of the four unique locus microsatellites tested, none yielded polymorphic products. Only two multi-locus microsatellites, having 15 and 17 consecutive repeats, respectively, exhibited polymorphism. Other microsatellites containing as many as 23 uninterrupted repeats were completely monomorphic. This has contrasted with Weber's (1990) findings, making it difficult to predict which microsatellites will be informative in white spruce.

Percentages of microsatellites associated with repetitive DNA have been found to vary between different species. Weber (1990) reported only 7% of human microsatellites to be embedded within repetitive DNA. This has contributed to the feasibility of using these markers in mammals. Estimates in plant species have ranged from 13.6% in cultivated potato (Provan *et al*, 1996) to 72.7% in *Arabidopsis* (Bell and Ecker, 1993). Approximately 25% of microsatellites in several forest tree species, including bur oak (Dow *et al*, 1995), eastern white pine (Echt *et al*, 1996), and Monterey pine (Smith and Devey, 1994) were found to be multi-locus. Six of the ten microsatellites isolated in white spruce, including both of the polymorphic microsatellites identified, were associated with repetitive DNA. This high frequency relative to other species suggests that further screening of potential microsatellite markers, prior to sequencing and design of primers, is advisable to eliminate those associated with repetitive DNA. Dot blot hybridizations of microsatellite loci with genomic DNA have been used with success for this purpose in Monterey pine (Smith and Devey, 1994).

No trinucleotide repeats were isolated in white spruce. Trinucleotide repeats are much easier to score than dinucleotide repeats (Kijas *et al*, 1995), but are also much more rare (Lagercrantz *et al*, 1993), and therefore much more expensive and time consuming to isolate. It is possible that hybridization conditions were too stringent for the $(TAT)_8$ and $(TCT)_8$ probes, which were hybridized together with the $(CT)_{12}$ probe. It is also likely that much more of the genome would need to be screened to locate these markers. As dinucleotide repeats are already relatively rare in white spruce, attempts to isolate trinucleotide repeats in this species would likely be prohibitively expensive, barring alternative techniques, such as successful enrichment of a library for trinucleotide repeats prior to screening (Ostrander, Jong, Rine *et al*, 1992).

Two variables, 5' to 3' DNA repair (Strand *et al*, 1993) and the temperature at which DNA extension occurs (Hite *et al*, 1996), have been implicated in slipped strand mis-pairing at microsatellite loci. It is possible that white spruce has particularily efficient DNA repair systems. This would be consistent with both reduced numbers of microsatellite loci, and reduced levels of polymorphism at these loci. For organisms having such long generation times, efficient DNA repair could be an important adaptation. Secondly, as a boreal forest species, white spruce occupies a colder climate than most plant species (Nienstaedt and Teich, 1972). Meiosis, in particular, occurs very early in spring, when temperatures are even further below average (Nienstaedt and Teich, 1972). Perhaps this contributes to stabilization of repeat tracts of DNA during replication.

4.4 Summary

Based upon the ten microsatellite loci which I characterized, microsatellites in white spruce appear to be much less common, as well as much less informative, than have been described in other species. In addition, a very high frequency of white spruce microsatellites appear to be associated with repetitive DNA, barring their usefulness in some applications. Of the ten loci identified, all of which exceeded ten uninterrupted tandem repeat units, none were polymorphic and single locus. Thus, it is difficult to predict how informative these markers might be in white spruce.

While no single locus, polymorphic microsatellites have yet been identified in white spruce, several have recently been reported in Sitka spruce (Van de Ven and McNicol, 1996). In this species, five of seven pairs of primers designed were found to amplify single locus, polymorphic products. Perfect repeat tracts having only six and eight units were informative. This makes it likely that sufficient numbers of informative microsatellite loci also exist within the white spruce genome.

The high cost of development of microsatellite markers in white spruce may be partially offset by the fact that relatively few polymorphic microsatellites will be required in order to complete population genetic surveys (Paetkau and Strobeck, 1994). If such markers are identified, population studies can be completed efficiently, partially because fewer markers are required for assessment than, for example with isozymes, and also because assessment involves use of the polymerase chain reaction, which allows large numbers of population samples to be processed quickly, using minimal amounts of sample DNA.

In spite of the extent to which white spruce is utilized in the Canadian forest sector, very few molecular genetic studies have been completed in this species. Development of a relatively small number of microsatellite loci in white spruce might allow data bases concerning genetic variation of this species to be compiled, as well as providing markers for such uses as paternity analysis, which is likely to be of increasing value to future tree improvement programs. It is uncertain at this time whether it will be cost effective to develop microsatellite loci at sufficient density to be useful for genetic mapping. Based on the results of my study, development of microsatellite markers for this purpose may be prohibitively expensive, barring improved techniques for their isolation. At this time, other marker systems, such as RFLP's or isozymes, appear to be more cost effective for this purpose.

Discussion Literature Cited

Alden, J. and C. Loopstra. 1987. Genetic diversity and population structure of *Picea glauca* on an altitudinal gradient in interior Alaska. Can. J. For. Res. **17**: 1519-1526.

Bell, C. J. and J. R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics, **19**: 137-144.

Bowers, J. E., G. S. Dangl, R. Vignani, and C. P. Meredith. 1996. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome, **39**: 628-633.

Cheliak, W. M., J. A. Pitel, and G. Murray. 1985. Population structure and the mating system of white spruce. Can. J. For. Res. **15**: 301-308.

Dow, B. D., M. V. Ashley and H. F. Howe. 1995. Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. Theor. Appl. Genet. **91**: 137-141.

Echt, C. S., P. May-Marquardt, M. Hseih, and R. Zahorchak. 1996. Characterization of microsatellite markers in eastern white pine. Genome, **39**: 1102-1108.

Hite, J. M., K. A. Eckert and K. C. Cheng. 1996. Factors affecting fidelity of DNA synthesis during PCR amplification of $d(C\text{-}A)_n \Diamond d(G\text{-}T)_n$ microsatellie repeats. Nucleic Acids Res. **24**: 2429-2434.

Kijas, J. M. H., J. C. S. Fowler and M. R. Thomas. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. Genome, **38**: 349-355.

Lagercrantz, U., H. Ellegren, and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**: 1111-1115.

Morgante, M., A. Rafalski, P. Biddle, S. Tingey and A. M. Olivieri. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome, **37**: 763-769.

Nienstaedt, H. and A. Teich. 1972. The genetics of white spruce. U.S.D.A. For. Serv. Res. Pap. WO-15. 24 pp.

Ostrander, E. A., P. M. Jong, J. Rine, and G. Duyk. 1992. Construction of small-insert genomic DNA libraries highly enriched for microstellite repeat sequences. Proc. Natl. Acad. Sci. USA. **89**: 3419-3423.

Paetkau, D, and C. Strobeck. 1994. Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology, **3**: 489-495.

Provan, J., W. Powell, and R. Waugh. 1996. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor. Appl. Genet. **92**: 1078-1084.

Smith D. N. and M. E. Devey. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome, **37**: 977-983.

Stoehr, M. U., and Y. A. El-Kassaby. 1997. Levels of genetic diversity at different stages of the domestication cycle of interior spruce in British Columbia. Theor. Appl. Genet. **94**: 83-90.

Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature (London), **356**: 274-276.

Taramino, G. and S. Tingey. 1996. Simply sequence repeats for germplasm analysis and mapping in maize. Genome, **39**: 277-287.

Terauchi, R. and A. Konuma. 1994. Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. Genome, **37**: 794-801.

Van de Ven, W. T. G., and R. J. McNicol. 1996. Microsatellites as DNA markers in sitka spruce. Theor Appl Genet. **93**: 613-617.

Weber, J. L. 1990. Informativeness of human $(dC\text{-}dA)_n$ / $(dG\text{-}dT)_n$ polymorphisms. Genomics, **7**: 524-530.

Wu, K. S. and S. D. Tanksley. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. **241**: 225-235.

Yazdani, R., O. Muona, D. Rudin, and A.E. Szmidt. 1985. Genetic structure of a *Pinus sylvestris* L. seed-tree stand and naturally regenerated understory. Forest Science, **31**: 430-436.